



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ**

**Εφαρμογή Προηγμένων Αλγορίθμων Επεξεργασίας  
Εικόνας για Εξαγωγή Αρχικών  
Αριθμητικών Δεδομένων απο  
Καμπύλες Επιβίωσης και Χρήση τους  
σε Μετα-ανάλυση**

**Σπύρος Ηλίας**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Υπεύθυνοι**

**Δρ.Π.Μπάγκος  
Δρ.Κ.Δελήμπασης  
Επίκουροι Καθηγητές**

**Λαμία, 2014**

## Ευχαριστίες

Η συγκεκριμένη εργασία έλαβε χώρα στη Λαμία στο Πανεπιστήμιο Θεσσαλίας του τμήματος Πληροφορικής με Εφαρμογές στην Βιοϊατρική από τον Φεβρουάριο του 2010 μέχρι και το Φεβρουάριο του 2014. Η παρούσα εργασία υλοποιήθηκε με την βοήθεια του εισηγητή του θέματος κ.Μπάγκου Παντελή σε συνεργασία με τον κ.Δελημπαση Κωνσταντίνο επίκουρους καθηγητές του τμήματος Πληροφορικής με Εφαρμογές στην Βιοϊατρική, τους οποίους ευχαριστώ για την βοήθεια τους. Θα ήθελα επίσης να ευχαριστήσω τους δικούς μου ανθρώπους για την υποστήριξη και την υπομονή τους.

## Περιεχόμενα

1	Εισαγωγή .....	6
1.1.1	Εισαγωγή στην Μετα-ανάλυση .....	6
1.1.2	Σχεδιασμός Μετα-ανάλυσης .....	7
1.1.3	Πλεονεκτήματα μειονεκτήματα .....	7
1.1.4	Επιστημονικά πεδία εφαρμογής της μετα-ανάλυσης .....	8
1.2	Τρόποι Πραγματοποίησης μετα-ανάλυσης και Στατιστικά Μεγέθη .....	9
1.2.1	Μέγεθος επίδρασης (effect size) .....	9
1.2.2	Μέγεθος επίδρασης Hazard Ratio .....	12
1.3	Ερμηνεία αποτελεσμάτων γραφήματος μετα-ανάλυσης .....	13
2	Μεθοδολογία .....	11
2.1	Σύγκριση fixed effect model και random effect model .....	14
2.2	Κλασικές Μέθοδοι μετα-ανάλυσης .....	14
2.3	Εναλλακτικές μέθοδοι μετα-ανάλυσης .....	15
2.4	Μεθοδολογία μετα-ανάλυσης από δημοσιευμένες καμπύλες Kaplan Meier .....	15
2.4.1	Καμπύλη επιβίωσης KAPLAN MEIER .....	16
2.5	Μέθοδος μετα-ανάλυσης από δημοσιευμένες καμπύλες .....	23
2.6	Προτεινόμενη μεθοδολογία και εξαγωγή δεδομένων από καμπύλες Kaplan Meier για την πραγματοποίηση μετα-ανάλυσης .....	26
2.6.1	Εισαγωγή εικόνας .....	27
2.6.2	Προεπεξεργασία του γραφήματος .....	27
2.6.3	Εντοπισμός καμπύλης .....	29
2.6.4	Έλεγχος εντοπισμού καμπύλης .....	31
2.6.5	Έλεγχος πλήθους καμπυλών .....	31
2.6.6	Καθορισμός κλίμακας αξόνων .....	32
2.6.7	Διαδικασία Μετα-ανάλυσης .....	33
3	Υλικά, υπολογιστική υλοποίηση .....	34
3.1	Διαθέσιμα δεδομένα .....	34
3.2	Περιβάλλον ανάπτυξης Matlab .....	34
3.3	Παρουσίαση της εφαρμογής .....	34
3.4	User interface .....	34
4	Αποτελέσματα .....	43
4.1	Εισαγωγή .....	43
4.2	Συνθετικές Καμπύλες .....	44
4.3	Αποτελέσματα από πραγματικές καμπύλες .....	53
4.3.1	Πραγματικό παράδειγμα Υπολογισμού του Log Hazard Ratio και της διακύμανσης από δημοσιευμένες καμπύλες .....	53
4.3.2	Παρουσίαση μελετών για καρκίνο στους πνεύμονες και μετα-ανάλυση .....	58
4.3.3	Μετα-ανάλυση .....	70
5	Συμπεράσματα .....	71
6	Βιβλιογραφία .....	73

## Περίληψη

Σε μελέτες που πραγματοποιούνται για την εύρεση νέων φαρμάκων ή θεραπειών διαφόρων νόσων και παθήσεων προκύπτουν σημαντικά αποτελέσματα. Τα αποτελέσματα αυτά από τις μεμονωμένες μελέτες επιβίωσης συνδυάζονται με στατιστικές μεθόδους ανάλυσης έτσι ώστε οι μελετητές να καταλήξουν σε ασφαλή συμπεράσματα όσον αφορά το αντικείμενο έρευνας. Η διαδικασία αυτή ονομάζεται μετα-ανάλυση. Οι μελέτες που χρησιμοποιούνται για την πραγματοποίηση μιας μετα-ανάλυσης ασχολούνται με το ίδιο αντικείμενο.

Τα δεδομένα των μελετών που συμπεριλαμβάνονται σε μια μετα-ανάλυση προέρχονται κυρίως από δημοσιευμένες μελέτες καθώς και από άλλες πηγές. Υπάρχουν πολλές μέθοδοι μετα-ανάλυσης οι οποίες για να εφαρμοστούν χρειάζονται τα ατομικά δεδομένα των συμμετεχόντων (individual patient data) από τις μεμονωμένες μελέτες. Συνήθως όμως δημοσιεύονται βασικά αποτελέσματα αυτών, όπως ένας εκτιμητής με τα διαστήματα εμπιστοσύνης του, ή μια καμπύλη επιβίωσης Kaplan-Meier και οι συμμετέχοντες της μελέτης. Οι κλασικές μέθοδοι εφαρμογής δεν είναι ικανές να χρησιμοποιηθούν για μετα-ανάλυση και έτσι προέκυψε η ανάγκη για την ανάπτυξη μιας μεθόδου που θα εκμεταλλεύεται τις πληροφορίες που δίνονται από τις καμπύλες επιβίωσης K-M.

Στην συγκεκριμένη εργασία παρουσιάζεται η ανάπτυξη μιας εφαρμογής που χρησιμοποιεί τα δεδομένα από τις καμπύλες K-M και εξάγει τις απαραίτητες πληροφορίες για την πραγματοποίηση μιας μετα-ανάλυσης. Για κάθε μία από τις επιμέρους μελέτες που θα συμπεριληφθούν στην μετα-ανάλυση η παρούσα εφαρμογή εκτελεί μια σειρά βημάτων μέχρι να προκύψει το τελικό αποτέλεσμα από την ολοκλήρωση της διαδικασίας. Τα βήματα αυτά είναι:

1. Εισαγωγή της εικόνας που περιέχει την καμπύλη επιβίωσης K-M και βασικές πληροφορίες της μελέτης.
2. Επεξεργασία της εικόνας για τον εντοπισμό των καμπυλών και εύρεση αριθμητικών στοιχείων της καμπύλης για την κάθε μελέτη.
3. Υπολογισμός στατιστικών μεγεθών για την κάθε καμπύλη και συνδυασμός αυτών για το αποτέλεσμα της μετα-ανάλυσης.

Η εφαρμογή αναπτύχθηκε σε περιβάλλον Matlab με την χρησιμοποίηση μιας σειράς αλγορίθμων και συναρτήσεων. Ο έλεγχος της ορθής λειτουργίας της εφαρμογής πραγματοποιήθηκε μέσα από αριθμητικές συγκρίσεις των αποτελεσμάτων της εφαρμογής με συγκεκριμένες καμπύλες K-M των οποίων τα δεδομένα ήταν ήδη γνωστά. Συγκεκριμένα τον υπολογισμό της διαφοράς μεταξύ των τιμών που προέκυψαν από την εφαρμογή και των πραγματικών τιμών των καμπυλών και την γραφική αναπαράσταση αυτών. Επίσης παρατίθενται και στοιχεία από την σύγκριση με εμπορικά διαθέσιμο λογισμικό αντίστοιχης λειτουργίας που δείχνει ότι η παρούσα υλοποίηση είναι πιο ολοκληρωμένη και απαιτεί λιγότερη αλληλεπίδραση με τον χρήστη.

Τέλος η εφαρμογή χρησιμοποιείται για την πραγματοποίηση μιας μετα-ανάλυσης σε 5 πραγματικές μελέτες οι οποίες προέρχονται από δημοσιεύσεις. Τα αποτελέσματα των μεμονωμένων μελετών καθώς και της συνολικής μετα-ανάλυσης που προέκυψαν από την χρήση της εφαρμογής συγκρίνονται με τα πραγματικά αποτελέσματα από τις δημοσιευμένες καμπύλες. Το τελικό αποτέλεσμα που προκύπτει από την λειτουργία της εφαρμογής συμπίπτει με αυτό της πραγματικής μελέτης.

**Λέξεις-Κλειδιά** μετα-ανάλυση, καμπύλες επιβίωσης Kaplan Meier, επεξεργασία εικόνας, μελέτες επιβίωσης

## Abstract

In this work we report the development of a computational tool, which estimates the numerical survival probabilities from bitmap K-M curves and performs meta-analysis of an arbitrary number of studies, using these obtained values. The tool uses image processing techniques and requires minimal user interaction. The proposed tool has been assessed quantitatively using the data from a published meta-analysis of five survival studies.

Meta-analysis of survival (time-to-event) data, from published studies is difficult, since there are different estimates that can be used and in many cases one cannot perform the transformation without having access to individual data. However, the majority of the published studies report a survival probability plot (K-M curve) that includes information for the survival probabilities as well as the censoring events. This work reports the development of a tool that performs meta-analysis using the bitmap of the individual K-M plots.

The proposed tool uses the bitmap of the K-M plots of any number of individual survival studies, as well as the total number of participants and minimum and maximum follow up for each study. A minimum user input is utilized to automatically track each survival curve and obtain its numeric values concerning time and survival probability. Dijkstra's algorithm was used for curve tracking and affine transformation was used to obtain the time and probability from the position of each detected pixel of the curve. The statistical quantities (the log-hazard ratio and its variance) are subsequently computed for each survival curve of each study. Finally, the log-hazard ratios are combined to calculate the pooled log hazard ratio. The proposed tool, developed using Matlab, is able to handle any number of survival studies, with intersecting K-M curves, with minimum user interaction: at least 2 point input for each curve in each K-M plot and at least 3 point input on the figure axes to define the scales of each axis.

First, results are presented concerning only the accuracy of obtaining the numerical values of the survival curves from the bitmaps of the K-M graphs. For this reason a number of synthetic K-M graphs were generated with known survival probability at different times. The proposed tool was used to extract the numeric values, which were compared to the ground truth. The same comparison took place using another, commercial WWW-based product. Results show that both methods work well and achieve median probability error below 1%. It has to be mentioned however that our approach operates with as many as 2 point inputs, whereas the approach in requires as many as 30 point inputs. Finally the proposed tool was used to repeat the meta-analysis of 5 studies investigating the benefit of postoperative cisplatin-based chemotherapy in patients with non-small-cell lung cancer. The pooled estimate of the log-hazard ratio and its 95% confidence interval that was extracted from the K-M curves was compared against the results reported in the published meta-analysis and it was found nearly identical (up to the third decimal place).

A computational tool has been reported that can perform meta-analysis of individual studies, using the bitmaps of K-M graphs. Initial results show the accuracy of the proposed tool.

**Keywords** meta-analysis, survival curves Kaplan Meier, image processing, survival studies

# 1 Εισαγωγή

## 1.1.1 Εισαγωγή στην Μετα-ανάλυση

Σημαντικές μελέτες επιβίωσης πραγματοποιούνται με σκοπό την ανακάλυψη νέων μεθόδων θεραπείας. Σε πολλές περιπτώσεις οι μεμονωμένες μελέτες δεν είναι ικανές να έχουν στατιστικά σημαντικό αποτέλεσμα ανεξάρτητα από το μέγεθος του δείγματος. Για το λόγο αυτό πραγματοποιούνται αρκετές μελέτες με το ίδιο αντικείμενο έρευνας. Οι μελέτες αυτές ξεχωριστά περιέχουν κάποια μικρή πληροφορία αλλά όλες μαζί μπορούν να δώσουν ένα σημαντικό αποτέλεσμα. Έτσι προέκυψε η ανάγκη να χρησιμοποιηθούν τυπικές στατιστικές μέθοδοι ελέγχου της ετερογένειας μέσω της σύνθεσης των μεμονωμένων αποτελεσμάτων με την διαδικασία της μετα-ανάλυσης [1]. Ο όρος μετα-ανάλυση [2] αναφέρεται στη στατιστική σύνθεση των αποτελεσμάτων από μια σειρά διαφορετικών πρωτογενών μελετών που ασχολούνται με το ίδιο αντικείμενο [3]. Πρόκειται για μια σχολαστική μελέτη που δεν εφαρμόζεται από μια μόνο μέθοδο αλλά ποικίλει ανάλογα με τα δεδομένα των μεμονωμένων μελετών. Στην ιατρική βιβλιογραφία, μετα-ανάλυση εφαρμόστηκε αρχικά στον τομέα των τυχαιοποιημένων κλινικών δοκιμών [4] αλλά σήμερα θεωρείται ένα πολύτιμο εργαλείο για τον συνδυασμό των αποτελεσμάτων από διαφορετικές μελέτες [5]. Με την πάροδο του χρόνου έχει εξελιχθεί και χρησιμοποιείται πλέον όχι μόνο στο χώρο της υγείας αλλά και σε άλλους τομείς.

Οι στατιστικές μέθοδοι που χρησιμοποιούνται σε μια μετα-ανάλυση μπορεί να εφαρμοστούν σε κάθε σύνολο δεδομένων, η σύνθεση θα έχει νόημα μόνο αν οι μελέτες έχουν συλλεχθεί συστηματικά. Η μέθοδος της μετα-ανάλυσης θεωρείται στατιστικά σημαντική καθώς συνδυάζει διαφορετικές μελέτες [6] που έχουν σχέση μεταξύ τους (ασχολούνται με το ίδιο αντικείμενο) αλλά είναι ταυτόχρονα και ανεξάρτητες. Οι μελέτες επιλέγονται με κατάλληλα με κριτήρια τα οποία καθορίζονται στο πρωτόκολλο της μετα-ανάλυσης.

Υπάρχουν διαφορετικές μέθοδοι πραγματοποίησης μιας μετα-ανάλυσης πάνω σε διάφορα σύνολα δεδομένων. Οι μέθοδοι μετα-ανάλυσης άρχισαν να αναπτύσσονται πριν από την δεκαετία του 90 και μέχρι σήμερα εξελίσσονται και βελτιώνονται. Η ιστορία ξεκινάει με την πραγματοποίηση των πρώτων προσπαθειών συνδυασμού και ομαδοποίησης αποτελεσμάτων μέσω αφηγηματικών ανασκοπήσεων για την εξαγωγή συμπερασμάτων. Οι αφηγηματικές ανασκοπήσεις [7] είναι μελέτες που γίνονται από ερευνητές χωρίς να υπάρχει σαφή διάκριση και κριτήρια επιλογής των μελετών που θα συμπεριληφθούν στην ανασκόπηση, ανομοιογένεια στον τρόπο λήψης πρωτογενών δεδομένων και άλλα σφάλματα που οδηγούν σε λανθασμένα συμπεράσματα. Με τη χρήση αυτής της τεχνικής υπάρχουν για παράδειγμα αναλύσεις για την αποτελεσματικότητα του ίδιου φαρμάκου που οι αναλυτές χρησιμοποιούσαν διαφορετικά όρια στην μελέτη με αποτέλεσμα να καταλήγουν σε αντίθετα συμπεράσματα. Για το λόγο αυτό πολύ σημαντικό ρόλο παίζει ο σχεδιασμός μιας μετα-ανάλυσης.

Για την πραγματοποίηση μιας ποιοτικής [8] και αποτελεσματικής μετα-ανάλυσης είναι πολύ σημαντικό το στάδιο αναζήτησης και εύρεσης των κατάλληλων μελετών. Αυτό θα επιτευχθεί με την μέθοδο της συστηματικής ανασκόπησης καθώς αποτελεί ουσιαστικά την καρδιά της μετα-ανάλυσης. Ειδικά στις περιπτώσεις μετα-αναλύσεων με αντικείμενο κλινικές μελέτες για κάποιο νέο φάρμακο ή θεραπεία, είναι ζωτικής σημασίας ο σωστός σχεδιασμός και η επιλογή των μελετών που θα συμμετάσχουν σε αυτή, αν αναλογιστεί κανείς την σημαντικότητα των αποτελεσμάτων. Για το λόγο αυτό η αναζήτηση δεδομένων που έχουν σχέση με το αντικείμενο μελέτης γίνεται με μεγάλη προσοχή και σε κάθε μέσο που μπορεί να παρέχει πληροφορία δηλαδή, σε βάσεις δεδομένων στο διαδίκτυο, σε περιοδικά, σε βιβλία και σε δημοσιεύτες μελέτες ή ακόμα και σε μελέτες που σε εξέλιξη.

Σκοπός της παρούσας εργασίας είναι η μετα-ανάλυση σε κλινικές μελέτες με την μέθοδο αξιοποίησης πληροφοριών από τα δεδομένα που παρέχουν οι καμπύλες επιβίωσης Kaplan Meier που θα παρουσιαστούν στο κεφάλαιο 2.

### 1.1.2 Σχεδιασμός Μετα-ανάλυσης

Για τον σχεδιασμό μιας μετα-ανάλυσης υπάρχουν κάποια βασικά βήματα που πρέπει να εκτελεσθούν όπως γίνεται σε κάθε είδους μελέτης ή έρευνας [9]. Το πρώτο βήμα περιλαμβάνει τον ορισμό του προβλήματος ή του αντικειμένου της έρευνας. Οι ερευνητές καλούνται δηλαδή να γράψουν ένα λεπτομερές πρωτόκολλο της μελέτης που δηλώνει τους στόχους, τις υποθέσεις που πρέπει να ελεγχθούν, και τα υποπεδία που μπορεί να σχετίζονται με το συγκεκριμένο θέμα. Στην συνέχεια ακολουθεί το βήμα των προτεινόμενων μεθόδων και κριτηρίων επιλογής σχετικών μελετών με το αντικείμενο της έρευνας. Η συμπερίληψη ή ο αποκλεισμός ορισμένων μελετών μπορεί να στρεβλώσει σε μεγάλο βαθμό τα αποτελέσματα της μετα-ανάλυσης. Ως εκ τούτου, τα κριτήρια για την επιλογή των μελετών πρέπει να καθορίζονται εκ των προτέρων με την βιβλιογραφική έρευνα [10]. Τέλος έχουμε το βήμα που περιλαμβάνει τις μεθόδους εξαγωγής δεδομένων και πληροφοριών για περαιτέρω επεξεργασία και οδήγηση σε σημαντικά στατιστικά συμπεράσματα. Αποτελέσματα που βασίζονται σε αποδείξεις και σημαντικά συμπεράσματα για μελλοντικές μελέτες.

Τα βήματα που παρουσιάστηκαν θα μπορούσαν να χωριστούν σε ακόμη πιο μικρά. Κάποια βασικά θέματα που πρέπει να δοθεί ιδιαίτερη προσοχή είναι:

1. Η αποφυγή συγχώνευσης διαφορετικών τύπων μελετών για να έχουμε ορθότερα συμπεράσματα.
2. Ο ορισμός αυστηρών κριτηρίων αποκλεισμού μελετών για την επίτευξη μεγαλύτερης ομοιογένειας.
3. Η αντιμετώπιση της κάθε μελέτης σαν μια ξεχωριστή οντότητα για καλύτερα αποτελέσματα.
4. Εκτίμηση του σφάλματος δημοσίευσης δηλαδή έλεγχος των μελετών που συμπεριλαμβάνουμε στην έρευνα γιατί αυτές που δημοσιεύονται πιο εύκολα γενικά είναι οι μελέτες με τα θετικά μόνο αποτελέσματα άρα οδηγούμαστε σε λάθος συμπεράσματα.

### 1.1.3 Πλεονεκτήματα και μειονεκτήματα

#### Πλεονεκτήματα

- Παρέχει μια αντικειμενική σύνθεση πολλών σχετικών μελετών ή πειραμάτων για ένα συγκεκριμένο πρόβλημα προκειμένου να οδηγηθεί η έρευνα σε ένα μοναδικό συνολικό συμπέρασμα.
- Η μετα-ανάλυση είναι ένας πολύ καλός τρόπος για την μείωση της πολυπλοκότητας, του μεγέθους της έρευνας και έχει σαν αποτέλεσμα την εξασφάλιση χρόνου και χρήματος βοηθώντας τους μελετητές να συλλέγουν πολύ περισσότερα δεδομένα χωρίς να απασχολείται μια ολόκληρη ερευνητική ομάδα.
- Η σωστή διαδικασία εφαρμογής της μετα-ανάλυσης επιτρέπει την αξιολόγηση της μεροληπτικότητας των μελετών που συμμετέχουν στην διαδικασία με την πραγματοποίηση στατιστικών ελέγχων που διασφαλίζουν την εγκυρότητα των αποτελεσμάτων.
- Η διαδικασία επιλογής των μελετών που θα συμμετάσχουν στην μετα-ανάλυση έχει γίνει πιο αυστηρή με καθορισμένα κάθε φορά κριτήρια και αυτό έχει ως αποτέλεσμα τα συμπεράσματα που προκύπτουν από την μετα-ανάλυση να είναι πολύ πιο αξιόπιστα.

#### Μειονεκτήματα

- Δεν υπάρχει κάποια συγκεκριμένη μεθοδολογία επιλογής των μελετών που θα συμμετάσχουν στην μετα-ανάλυση παρά μόνο η στρατηγική αναζήτησης και επιλογής που θα εφαρμοστεί από τον κάθε ερευνητή με κριτήρια που θα θεωρήσει ο ίδιος κατάλληλα.
- Η μετα-ανάλυση είναι πολύ συχνά μεροληπτική προς την κατεύθυνση ενός γενικού συμπεράσματος στατιστικής σημαντικότητας, επειδή δεν έχει την δυνατότητα να πάρει υπόψη της τα αποτελέσματα μελετών που έχουν οδηγήσει σε μη σημαντικά στατιστικά συμπεράσματα διότι δεν έχουν δημοσιευθεί.
- Η μετα-ανάλυση συχνά οδηγεί σε μία υπεραπλούστευση κάποιων ερευνητικών προβλημάτων δεδομένου ότι επικεντρώνει την προσοχή της σε μια συνολική μέση επίδραση αντί να μελετήσει τους ενδιάμεσους παράγοντες.

#### 1.1.4 Επιστημονικά πεδία εφαρμογής της μετα-ανάλυσης

Στην ιατρική, οι συστηματικές ανασκοπήσεις και η μετα-ανάλυση αποτελούν τον πυρήνα του κινήματος για την εξασφάλιση ότι οι ιατρικές θεραπείες βασίζονται στα βέλτιστα διαθέσιμα εμπειρικά δεδομένα. Για παράδειγμα, το Cochrane Collaboration [11] έχει δημοσιεύσει αποτελέσματα για περισσότερες από 3700 μετα-αναλύσεις (Από τον Ιανουάριο του 2009) τα οποία συνθέτουν τα δεδομένα σχετικά με τις θεραπείες σε όλους τους τομείς της υγείας, συμπεριλαμβανομένων και των ειδικών περιπτώσεων των πονοκέφαλων, του καρκίνου, των αλλεργιών, των καρδιαγγειακών παθήσεων, την πρόληψη του πόνου, και την κατάθλιψη. Πραγματοποιείτε επίσης έλεγχος σε στοιχεία που έχουν σχέση με παρεμβάσεις που σχετίζονται με νέο-μεταγεννητική φροντίδα, τον τοκετό, βρεφικές και παιδικές ασθένειες, καθώς και ασθένειες κοινές σε εφήβους, ενήλικες και ηλικιωμένους. Τα είδη των παρεμβάσεων περιλαμβάνουν χειρουργική επέμβαση, τα φάρμακα, τον βελονισμό και κοινωνικές παρεμβάσεις. Το περιοδικό BMJ [12] δημοσιεύει μια σειρά από άρθρα με Evidence Based Medicine, που βασίστηκε στα αποτελέσματα συστηματικών ανασκοπήσεων. Οι συστηματικές ανασκοπήσεις και οι μετα-αναλύσεις χρησιμοποιούνται επίσης για την μέτρηση της απόδοσης διαφόρων διαγνωστικών εξετάσεων, και επιδημιολογικών συσχετίσεων μεταξύ της έκθεσης και ο επιπολασμού της νόσου, μεταξύ άλλων θεμάτων.

Ερευνητικά κέντρα [10] διεξάγουν συνήθως μια σειρά μελετών για την αξιολόγηση της αποτελεσματικότητας ενός φαρμάκου. Χρησιμοποιούν μετα-ανάλυση για να συνθέσουν τα δεδομένα από τις μελέτες αυτές, πραγματοποιώντας ένα πιο ισχυρό έλεγχο (και ακριβέστερη εκτίμηση) της επίδρασης του φαρμάκου. Επιπλέον, η μετα-ανάλυση παρέχει ένα πλαίσιο για την αξιολόγηση της σειράς μελετών ως σύνολο, αντί να ελέγχουν κάθε μια μεμονωμένα. Οι αναλύσεις αυτές παίζουν σημαντικό ρόλο στην εσωτερική έρευνα, για τις αναφορές στις κυβερνητικές υπηρεσίες, καθώς και στο μάρκετινγκ. Οι μετα-αναλύσεις χρησιμοποιούνται επίσης για να συνθέσουν τα στοιχεία σχετικά με ανεπιθύμητες παρενέργειες των φαρμάκων, δεδομένου ότι αυτά εκδηλώνονται σπάνια και θα πρέπει να συσσωρεύονται πληροφορίες σε μια σειρά μελετών για να εκτιμηθεί σωστά ο κίνδυνος αυτών των γεγονότων.

Στον τομέα της εκπαίδευσης η μετα-ανάλυση έχει εφαρμοστεί σε ποικίλα θέματα, όπως η σύγκριση της εκπαίδευσης εξ αποστάσεως με την παραδοσιακή διδασκαλία σε μια αίθουσα, η αξιολόγηση των επιπτώσεων της εκπαίδευσης στις αναπτυσσόμενες οικονομίες, διάφορα τεστ γνώσεων και αξιολόγησης καθώς επίσης και άλλα παιδαγωγικά πειράματα. Η μέθοδος της μετα-ανάλυσης [13] είναι πολύ σημαντική καθώς υπάρχει η δυνατότητα αξιοποίησης πληροφοριών από τις ποικίλες καταστάσεις και τους παράγοντες που επηρεάζουν την εκπαίδευση για την εξαγωγή σημαντικών συμπερασμάτων. Αποτελέσματα διαφόρων ερευνών και η χρήση αυτών σε μετααναλύσεις έχουν επηρεάσει σε πολύ μεγάλο βαθμό τον τρόπο εκπαίδευσης σε αρκετές περιοχές ανά τον κόσμο.

Στην ψυχολογία, η μετα-ανάλυση έχει εφαρμοστεί για την βασική επιστήμη, όπως επίσης και για την υποστήριξη θεωριών [14] όσον αφορά τον τρόπο λειτουργίας του ατόμου σε διάφορες καταστάσεις ανάλογα με την ψυχολογική του κατάσταση. Έχει χρησιμοποιηθεί για την αξιολόγηση στις αλλαγές της προσωπικότητας όλη τη διάρκεια της ζωής ενός ατόμου, για την εκτίμηση της επίδρασης της βίας μέσα στην επιθετική συμπεριφορά, την απόδοση και την στάση των ανθρώπων στο χώρο εργασίας με βάση τις πέντε διαστάσεις της συμπεριφοράς. Έχουν επίσης πραγματοποιηθεί πολύ σημαντικές έρευνες [2] για να γίνει εξέταση των διαφορών στα δύο φύλα στα πεδία της μαθηματικής ικανότητας, της ηγεσίας, και της μη λεκτικής επικοινωνίας. Αποτελεί πλέον ένα σύγχρονο εργαλείο στην επιλογή θεραπειών για ψυχολογικά προβλήματα, συμπεριλαμβανομένου της ψυχαναγκαστικής διαταραχής, της διαταραχής παρορμητικότητας, της νευρικής βουλιμίας, της κατάθλιψη, των φοβιών, και την διαταραχή πανικού.

Στον τομέα της εγκληματολογίας, κυβερνητικοί οργανισμοί έχουν χρηματοδοτήσει μετα-αναλύσεις που εξετάζουν τη σχετική αποτελεσματικότητα των διαφόρων προγραμμάτων για τη μείωση της εγκληματικής συμπεριφοράς. Αυτές περιλαμβάνουν πρωτοβουλίες για την πρόληψη της εγκληματικότητας, τη μείωση της υποτροπής, την αξιολόγηση της αποτελεσματικότητας των διαφορετικών στρατηγικών για περιπολίες της αστυνομίας και για τη χρήση των ειδικών δικαστηρίων που ασχολούνται με εγκλήματα που έχουν σχέση με ναρκωτικά. Πολλές μελέτες οδήγησαν [15] στο συμπέρασμα ότι δεν αρκούν οι ποινικές κυρώσεις χωρίς σωστή παροχή υπηρεσιών αποκατάστασης και χωρίς αναφορά σε ψυχολογικά αίτια έτσι προέκυψαν διάφορα μοντέλα αντιμετώπισης και βελτίωσης του σωφρονισμού των εγκληματιών .



Στον τομέα της οικολογίας, οι μετα-αναλύσεις που χρησιμοποιούνται για τον προσδιορισμό των περιβαλλοντικών επιπτώσεων των αιολικών πάρκων, βιοτικών αντιστάσεων στην εισβολή εξωτικών φυτών, οι επιπτώσεις των αλλαγών στην τροφική αλυσίδα των θαλασσών, οι αντιδράσεις των φυτών στην παγκόσμια κλιματική αλλαγή κ.α.

## 1.2 Τρόποι Πραγματοποίησης μετα-ανάλυσης και Στατιστικά Μεγέθη

Οι μετα-αναλύσεις γίνονται όλο και περισσότερο αποδεκτές ως το καλύτερο μέσο για να συνοψίζεις τα αποτελέσματα μιας σειράς τυχαιοποιημένων ελεγχόμενων μελετών ώστε να αντιμετωπίζονται παρόμοιες περιπτώσεις [16]. Ο καλύτερος τρόπος για να πραγματοποιηθεί μια μετα-ανάλυση σε όλες τις σχετικές τυχαιοποιημένες κλινικές μελέτες είναι η ανάλυση των στατιστικών δεδομένων από όλους τους ασθενείς που συμμετείχαν σε αυτές. Τα στατιστικά δεδομένα προκύπτουν από τα γεγονότα (time to event data) ή αλλιώς θανάτους των ασθενών κατά την διάρκεια μιας χρονικής περιόδου της μελέτης [17]. Τον πιο δύσκολο ρόλο στην όλη διαδικασία τον έχουν αυτοί που θα αναλάβουν την συλλογή τον έλεγχο και την ανάλυση των πρώτων δεδομένων από αυτές τις μελέτες με λίγα λόγια την διαδικασία της συστηματικής ανασκόπησης. Με αυτό τον τρόπο αυξάνεται η ακρίβεια και η δυνατότητα ενημέρωσης των δεδομένων καθώς επίσης η ευελιξία και η έκταση των αναλύσεων στην εκάστοτε μελέτη. Λιγότερο αυστηρές μετα-αναλύσεις (αφηγηματικές ανασκοπήσεις) που βασίζονται μόνο σε στοιχεία που προέρχονται από τις δημοσιευμένες εκθέσεις μπορεί να δώσουν διαφορετικά αποτελέσματα. Η συλλογή δεδομένων των επιμέρους ασθενών απαιτεί μεγάλο χρονικό διάστημα, αρκετούς πόρους και επιμονή. Ωστόσο μέχρι σήμερα αυτή η προσέγγιση έχει υιοθετηθεί σε ένα σχετικά μικρό αριθμό μετα-αναλύσεων. Για το λόγο αυτό, οι περισσότερες μετα-αναλύσεις βασίζονται στην άντληση δεδομένων από τη δημοσιευμένη βιβλιογραφία, διαδικτυακές βάσεις δεδομένων δημοσιεύσεις σε περιοδικά αλλά και μελέτες που δεν έχουν δημοσιευτεί. Έτσι μπορούμε να έχουμε την δυνατότητα να επεξεργαστούμε μεγαλύτερο όγκο δεδομένων σε μικρότερο χρονικό διάστημα χρησιμοποιώντας λιγότερους πόρους. Γενικά υπάρχουν διαφορετικοί τρόποι και στατιστικές μεταβλητές για την πραγματοποίηση μιας μετα-ανάλυσης.

### 1.2.1 Μέγεθος επίδρασης (effect size)

Οι μεταβλητές που χρησιμοποιούνται για την πραγματοποίηση μιας μετα-ανάλυσης περιγράφονται από κάποιο μέγεθος επίδρασης (effect size), που ουσιαστικά είναι μια τιμή η οποία αντικατοπτρίζει το ποσοστό επίδρασης της θεραπείας (treatment effect) σε διάφορες κλινικές μελέτες ή την σχέση μεταξύ δύο μεταβλητών για πολλών ειδών μελέτες [2]. Υπολογίζεται αρχικά το μέγεθος επίδρασης για κάθε μια από τις μελέτες που θα συμπεριληφθούν στην συνολική μετα-ανάλυση και έπειτα χρησιμοποιούμε τα επιμέρους αποτελέσματα για τον υπολογισμό της συνολικής επίδρασης για να καταλήξουμε σε χρήσιμα στατιστικά συμπεράσματα. Οι μετα-αναλύσεις σε κλινικές μελέτες για να περιγράψουν την επίδραση της θεραπείας χρησιμοποιούν διάφορα μεγέθη επίδρασης όπως Odds Ratio , Relative Risk, Risk Differences.

Πολύ σημαντικό πριν την πραγματοποίηση μιας μετα-ανάλυσης είναι η επιλογή του κατάλληλου μεγέθους επίδρασης. Τα κριτήρια για αυτήν την επιλογή είναι:

- Οι διαφορετικές μελέτες που θα συμπεριληφθούν στην μετα-ανάλυση να περιέχουν συγκρίσιμα μεταξύ τους μεγέθη στα αποτελέσματα τους.
- Στα στοιχεία που παρατίθενται στην εκάστοτε μελέτη πρέπει να υπάρχουν τα αποτελέσματα από τις διάφορες μετρήσεις που θα χρησιμοποιηθούν για την πραγματοποίηση της μετα-ανάλυσης.
- Το μέγεθος επίδρασης να έχει καλές τεχνικές προδιαγραφές. Για παράδειγμα πρέπει να δίνονται στοιχεία όπως η κατανομή του δείγματος και η διακύμανση του (ή να μπορούν να υπολογιστούν) για την πραγματοποίηση της μελέτης.

Ο Σχετικός Κίνδυνος (Relative Risk) είναι ο λόγος των κινδύνων(risks) των δύο πληθυσμών [6] όπως ορίζεται στη σχέση 1. Για παράδειγμα εξετάζουμε τους 2 πληθυσμούς που παρουσιάζονται στον πίνακα ενδεχομένων 2. Οι πληθυσμοί χωρίζονται σε αυτούς που ακολουθούν την κλασική θεραπεία (control group) που είναι συνολικά 100 άτομα και αυτούς που ακολουθούν την νέα μέθοδο θεραπείας (research group) που είναι επίσης 100 στο σύνολο. Έστω ότι πεθαίνουν 5 από το research group και 10 από το control group, προκύπτει έτσι ένας σχετικός κίνδυνος στο research group που ισούται με 5/100 και ένας σχετικός κίνδυνος για το control group που ισούται με 10/100 έτσι ο λόγος των δύο κινδύνων είναι 0,5 όπως προκύπτει από τη

σχέση 1. Για το Relative Risk οι υπολογισμοί συνεχίζονται στην λογαριθμική κλίμακα. Συγκεκριμένα αφού υπολογιστεί ο φυσικός λογάριθμος του Relative Risk (σχέση 2), η διακύμανση(σχέση 3), το σχετικό του σφάλμα (σχέση 4) και το διάστημα εμπιστοσύνης [(σχέση 6,7) LL, UL] γίνεται η χρήση αυτών των αριθμών για την πραγματοποίηση της μετα-ανάλυσης. Όταν ολοκληρωθεί η όλη διαδικασία της μετα-ανάλυσης, μετατρέπουμε τα αποτελέσματα στην αρχική τους κλίμακα [2]. Η εικόνα 1 είναι μια γραφική αναπαράσταση μετα-ανάλυσης τριών μελετών με 2 πληθυσμούς ανά μελέτη που χρησιμοποιεί το Relative Risk. Τα διαστήματα εμπιστοσύνης που αναφέρθηκαν παραπάνω περιγράφουν την ακρίβεια με την οποία έχει υπολογιστεί το εκάστοτε Effect Size για την κάθε μελέτη που συμμετέχει σε μια μετα-ανάλυση. Η μετατροπή σε λογαριθμική κλίμακα είναι απαραίτητη για να διατηρηθεί η συμμετρία.

Πίνακας 1 Παράδειγμα πίνακα με γενικά αποτελέσματα μελέτης.

	Γεγονότα	Μη-Γεγονότα	N
Research Group	A	B	N1
Control Group	C	D	N2

Πίνακας 2 Συνθετικός πίνακας με τυχαίους αριθμούς που αναπαριστά μια μελέτη.

	Νεκροί	Ζωντανοί	Ασθενείς
Research Group	5	95	100
Control Group	10	90	100

$$RelativeRisk = \frac{A/N1}{C/N2} \quad (1)$$

$$LogRelativeRisk = \ln(RelativeRisk) \quad (2)$$

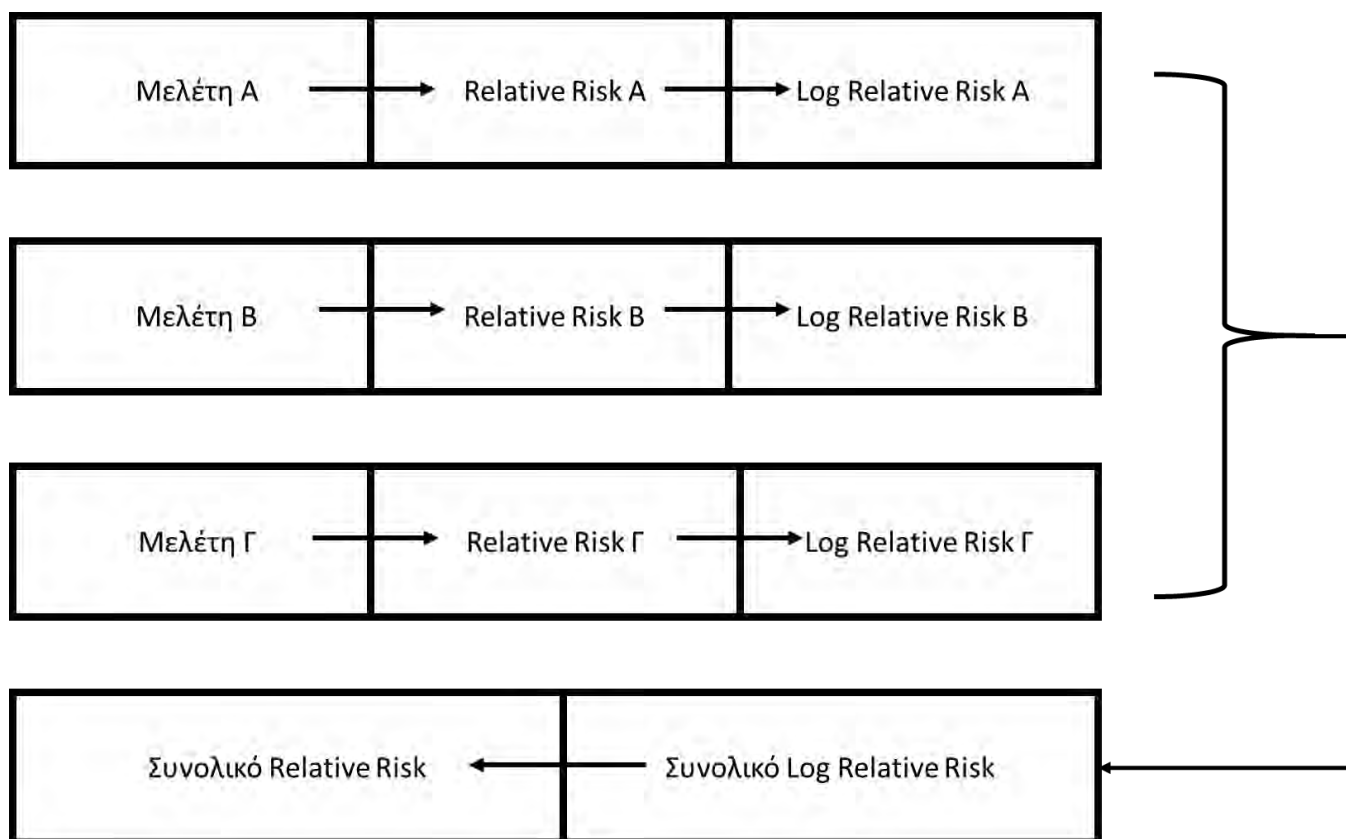
$$V_{LogRelativeRisk} = \frac{1}{A} - \frac{1}{N1} + \frac{1}{C} - \frac{1}{N2} \quad (3)$$

$$SE_{LogRelativeRisk} = \sqrt{V_{LogRelativeRisk}} \quad (4)$$

$$RelativeRisk = \exp(LogRelativeRisk) \quad (5)$$

$$LL_{RelativeRisk} = \exp(LL_{LogRelativeRisk}) \quad (6)$$

$$UL_{RelativeRisk} = \exp(UL_{LogRelativeRisk}) \quad (7)$$



Εικόνα 1 Διαδικασία μετα-ανάλυσης με μέγεθος επίδρασης το Relative Risk και κάθε μελέτη να έχει 2 πληθυσμούς [2].

Ένα άλλο effect size είναι το Odds Ratio που υπολογίζει την πιθανότητα γεγονότος (θανάτου) σε σχέση με την πιθανότητα μη γεγονότος. Για παράδειγμα στην περίπτωση που χρησιμοποιηθεί σαν Effect Size το Odds Ratio και με βάση τον πίνακα 2 προκύπτει για το control group μια πιθανότητα θανάτου 5/95 (0.0526) και για το research group μια πιθανότητα 10/90 (0.1111). Το Relative Risk θεωρείται από αρκετούς καλύτερο Effect Size από ότι το Odds Ratio αλλά οι στατιστικές δυνατότητες που προσφέρει το δεύτερο το καθιστούν καλύτερη επιλογή για την πραγματοποίηση μιας μετα-ανάλυσης. Όταν ο αριθμός των θανάτων είναι μικρός το Odds Ratio θα είναι παρεμφερές με το Relative Risk. Στην περίπτωση επιλογής του Odds Ratio επίσης αλλάζουμε την κλίμακα σε λογαριθμική για τον ίδιο λόγο που εφαρμόστηκε και στο Relative Risk. Η διαδικασία της μετα-ανάλυσης δηλαδή περιλαμβάνει υπολογισμό του Odds Ratio (σχέση 8), μετατροπή στην λογαριθμική κλίμακα (σχέση 9), υπολογισμό της διακύμανσης (σχέση 10), του τυπικού σφάλματος και των διαστημάτων εμπιστοσύνης και μετατροπή στην αρχική κλίμακα (οι τελευταίες μεταβλητές υπολογίζονται όπως οι αντίστοιχες του Relative Risk). Οι σχέσεις που θα χρησιμοποιηθούν είναι παρόμοιες με αυτές του Relative Risk.

$$OddsRatio = \frac{A \cdot D}{B \cdot C} \quad (8)$$

$$LogOddsRatio = \ln(OddsRatio) \quad (9)$$

$$V_{OddsRatio} = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D} \quad (10)$$

,όπου A, B, C, D όπως έχουν οριστεί στον πίνακα 1. Ένα ακόμη σημαντικό Effect Size που πρέπει να αναφερθεί είναι το Risk Difference (διαφορά κινδύνου) που είναι ουσιαστικά η διαφορά μεταξύ των δύο κινδύνων. Με

βάση τα παραπάνω στοιχεία του πίνακα 1 με κίνδυνο 0.05 για το research group και 0.10 για το control group προκύπτει ένα Risk Difference που ισούται με -0,05. Βασική διαφορά από τα δυο προηγούμενα Effect Sizes είναι ότι όλες οι πράξεις γίνονται χωρίς να αλλάξει κλίμακα η μεταβλητή μας. Η διαδικασία μετα-ανάλυσης με το συγκεκριμένο Effect Size περιλαμβάνει τον υπολογισμό του Risk Difference, της διακύμανσης και του τυπικού σφάλματος.

### 1.2.2 Μέγεθος επίδρασης Hazard Ratio

Στην περίπτωση της παρούσας εργασίας το Effect Size που θα χρησιμοποιήσουμε είναι παρόμοιο με το Relative Risk και είναι το Hazard Ratio (HR). Δηλαδή τα στατιστικά αποτελέσματα που είναι χρήσιμα να εξαχθούν από την ελεγχόμενη τυχαιοποιημένη μελέτη επιβίωσης ασθενών είναι το Hazard Ratio και η διακύμανση. Συγκεκριμένα το Hazard Ratio [16] ερμηνεύει την αποτελεσματικότητα κάποιας νέας θεραπείας ή φαρμάκου με την σύγκριση δύο καμπυλών επιβίωσης που αναπαριστούν 2 διαφορετικούς πληθυσμούς. Ο ένας πληθυσμός είναι αυτός που του χορηγείται ένα νέο φάρμακο ή κάποια νέα θεραπεία και ο άλλος πληθυσμός είναι αυτός στον οποίο εφαρμόζονται κλασικές μέθοδοι θεραπείας (Research group και Control group). Το Hazard Ratio είναι η μοναδική στατιστική μεταβλητή που συνοψίζει το censoring και τον χρόνο μέχρι την εμφάνιση κάποιου γεγονότος (time to event). Ο όρος 'censoring' αναφέρεται σε άτομα που σταμάτησαν να συμμετέχουν στην μελέτη χωρίς να είναι γνωστό σε τι κατάσταση βρίσκονται. Το 'γεγονός' είναι ο θάνατος κάποιου ατόμου κατά την διάρκεια συμμετοχής του στη μελέτη. Αν δεν λαμβάνεται υπόψιν το censoring και ο χρόνος μέχρι κάποιο γεγονός στον υπολογισμό του log Hazard Ratio τότε αυτό που προκύπτει είναι το relative risk.

Συχνά αναφέρεται ότι η χρήση του Hazard Ratio έμμεσα υποθέτει ανάλογους κινδύνους και για τους δύο πληθυσμούς που συγκρίνονται, αλλά αυτό δεν είναι απόλυτο για κάθε περίπτωση. Το Hazard Ratio είναι μια μεταβλητή που συνοψίζει την διαφορά των καμπυλών επιβίωσης δυο πληθυσμών και αναπαριστά την συνολική μείωση στον κίνδυνο θανάτου από μια συγκεκριμένη θεραπεία σε σχέση με τον έλεγχο καθ' όλη τη διάρκεια της παρακολούθησης του ασθενούς. Για να υπάρχει η δυνατότητα καλύτερης ερμηνείας αυτού του στατιστικού μεγέθους είναι πολύ σημαντική η γραφική αναπαράσταση του είτε με την μορφή καμπύλης επιβίωσης είτε με την μορφή αναπαράστασης γραφήματος κινδύνου (hazard function plot). Οι καμπύλες επιβίωσης παρουσιάζονται αναλυτικά στο κεφάλαιο 2.

Θεωρούμε ότι σε μια μετα-ανάλυση συμμετέχουν  $K$  μελέτες και για κάθε  $i=1, \dots, K$  μελέτη θα υπολογίζεται Log Hazard Ratio  $\ln(HR_i)$  και η διακύμανση της εκάστοτε μελέτης  $var[\ln(HR_i)]$  ξεχωριστά. Το Hazard ratio ορίζεται με την βοήθεια του relative risk από την σχέση  $Hr = (RR - 1)/RR$ . Περισσότερες λεπτομέρειες θα αναφερθούν στην μεθοδολογία που ακολουθεί σε παρακάτω κεφάλαιο. Αφού συγκεντρωθούν τα αποτελέσματα αυτά θα υπολογιστεί τελικά ένα συνολικό log Hazard ratio  $\ln(HR)$  από την σχέση 11 και μια συνολική διακύμανση  $var[HR]$  από την σχέση 12 για τις  $K$  μελέτες.

$$\ln(HR) = \frac{\sum_{i=1}^K \frac{\ln(HR_i)}{var[HR_i]}}{\sum_{i=1}^K \frac{1}{var[HR_i]}} \quad (11)$$

$$var[\ln(HR)] = \left[ \sum_{i=1}^K \frac{1}{var[HR_i]} \right] \quad (12)$$

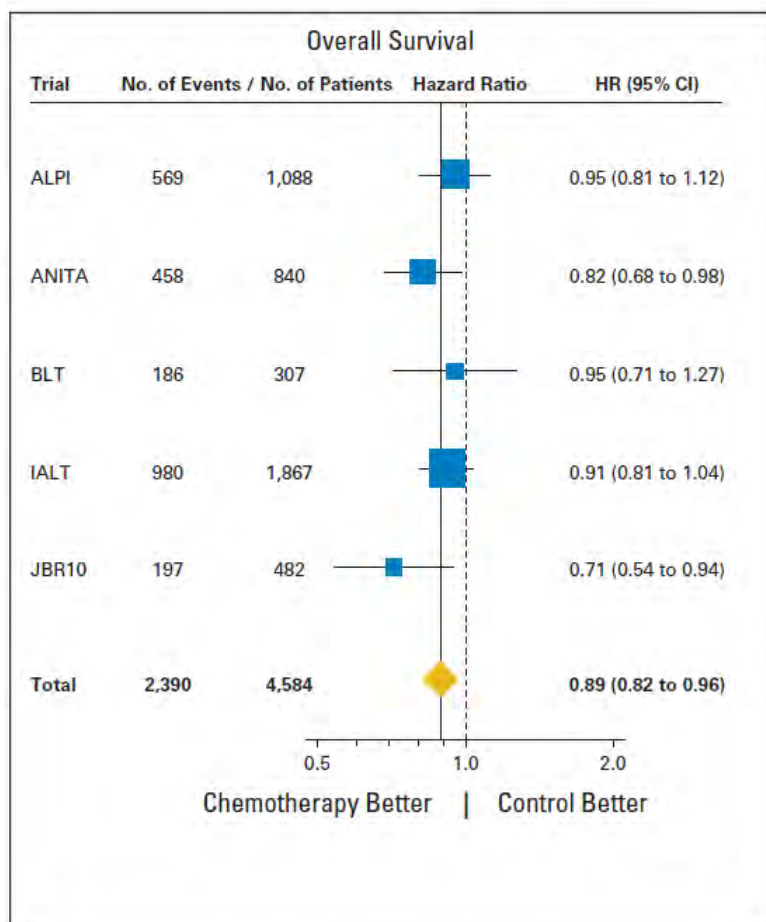
Είναι προφανές από τους τύπους ότι η συνολική πιθανότητα κινδύνου είναι ένας σταθμισμένος μέσος όρος των επιμέρους πιθανοτήτων κινδύνου από την κάθε μελέτη με τα βάρη αντιστρόφως ανάλογα προς την διακύμανση του log hazard για κάθε δοκιμή. Η μέθοδος υπολογισμού αυτών των δύο στατιστικών μεγεθών θα παρουσιαστεί με λεπτομέρειες παρακάτω.

### 1.3 Ερμηνεία αποτελεσμάτων γραφήματος μετα-ανάλυσης

Τα αποτελέσματα μιας ολοκληρωμένης μετα-ανάλυσης έχουν ένα συγκεκριμένο τρόπο γραφικής απεικόνισης όπως την εικόνα 2 που ονομάζεται forest plot. Ο πίνακας που αναπαριστά ουσιαστικά η εικόνα 2 αποτελείται από:

- Την 1<sup>η</sup> στήλη που περιέχει το όνομα της μελέτης
- Την 2<sup>η</sup> στήλη που περιέχει το σύνολο των γεγονότων (πλήθος θανάτων ανά μελέτη)
- Την 3<sup>η</sup> στήλη που περιέχει το σύνολο των συμμετεχόντων στην εκάστοτε μελέτη
- Την 4<sup>η</sup> στήλη που περιέχει την γραφική αναπαράσταση του αποτελέσματος της μετα-ανάλυσης
- Την 5<sup>η</sup> στήλη που περιέχει τα αποτελέσματα του διαστήματος εμπιστοσύνης με ακρίβεια 95%

Το κάθε τετράγωνο στην 4<sup>η</sup> στήλη αναπαριστά ουσιαστικά το αποτέλεσμα της εκάστοτε μελέτης με το μέγεθος να διαφέρει ανάλογα με την βαρύτητα που δίνεται στις διαφορετικές μελέτες όταν υπολογίζεται η συνολική επίδραση. Το μέγεθός του δηλαδή επηρεάζεται από το σύνολο των συμμετεχόντων στην εκάστοτε μελέτη (μέγεθος του δείγματος). Έτσι στις μελέτες τις οποίες συμμετείχαν περισσότεροι ασθενείς δίνονταν μεγαλύτερη βαρύτητα και αντίστοιχα λιγότερη σε αυτές με λιγότερους ασθενείς. Οι οριζόντιες γραμμές που ενώνονται με τα τετράγωνα αναπαριστούν το διάστημα εμπιστοσύνης 95%. Το διαμάντι στην εικόνα 2 αναπαριστά το συνολικό αποτέλεσμα της μετα-ανάλυσης που προκύπτει από όλες τις μελέτες [2]. Ο οριζόντιος άξονας με τις υποδιαίρέσεις, αναπαριστά τις τιμές του Hazard Ratio ανάλογα με τα αποτελέσματα από την μετα-ανάλυση. Αν το 1 βρίσκεται στο 95 % διάστημα εμπιστοσύνης σημαίνει ότι δεν υπάρχει καμία στατιστική σημαντικότητα στο 5%. Αν ένα τετράγωνο βρίσκεται στη μέση της κάθετης γραμμής που ξεκινάει από το 1 σημαίνει δεν υπάρχει κάποια διαφορά στην θεραπεία που χρησιμοποιήθηκε στο control group σε σχέση με αυτή του research group [17]. Αν ένα τετράγωνο βρίσκεται δεξιά της κάθετης γραμμής που είναι ένα σημαίνει ότι η θεραπεία του control group (παλιός τρόπος θεραπείας) είναι αποτελεσματικότερη της θεραπείας του research group (θεραπεία υπό έρευνα) και αν είναι στα αριστερά σημαίνει ακριβώς το αντίθετο. Στην τελευταία στήλη έχουμε το πραγματικό αποτέλεσμα της εκάστοτε μελέτης και το διάστημα εμπιστοσύνης στην παρένθεση.



Εικόνα 2 Αποτελέσματα των μελετών που συμμετείχαν στην μετα-ανάλυση και συνολικό αποτέλεσμα (forest plot) [18] .

## 2 Μεθοδολογία

Τα δεδομένα που χρησιμοποιούνται για την πραγματοποίηση μιας μετα-ανάλυσης τις περισσότερες φορές αντλούνται από δημοσιευμένες μελέτες. Η μετα-ανάλυση μπορεί να εφαρμοστεί με διαφορετικές μεθόδους ανάλογα με τα δεδομένα που υπάρχουν στη διάθεση του αναλυτή. Οι μέθοδοι χωρίζονται σε αυτές που υπολογίζουν fixed effect model, αυτές που υπολογίζουν random effect model, η και τα δυο [2]. Η μέθοδος που θα χρησιμοποιηθεί στην παρούσα εργασία βασίζεται σε καμπύλες επιβίωσης Kaplan-Meier και βασικά στοιχεία της μελέτης όπως θα παρουσιαστούν παρακάτω. Για την καλύτερη αναγνώριση και κατανόηση των διαφορετικών μεθόδων παρουσιάζεται μια άτυπη διαίρεση αυτών σε κλασικές και εναλλακτικές μεθόδους.

### 2.1 Σύγκριση fixed effect model και random effect model

Το fixed effect model αναφέρεται σε effect sizes που ισχύει σαν αρχική υπόθεση το γεγονός ότι οι πληθυσμοί που συμμετέχουν στις μελέτες είναι πανομοιότυποι, δεν υπάρχουν παράγοντες που να τους κάνουν να διαφέρουν ή αυτοί ο παράγοντας είναι ίδιοι για όλους τους συμμετέχοντες [19]. Αυτό σημαίνει ότι το πραγματικό effect size που υπολογίζεται έχει ίδιο αποτέλεσμα σε όλες τις μελέτες. Εφόσον χρησιμοποιείται το ίδιο πραγματικό effect size για όλες τις μελέτες, το υπολογισμένο effect size ποικίλει επειδή υπάρχει το τυχαίο λάθος που εμφανίζεται σε κάθε μελέτη. Αυτό το λάθος δεν θα υπήρχε μόνο στην περίπτωση που το μέγεθος του δείγματος μας δηλαδή οι συμμετέχοντες θα ήταν άπειροι.

Το random effect model αναφέρεται σε μελέτες που πραγματοποιούνται και λαμβάνονται υπόψιν τα διαφορετικά χαρακτηριστικά ή οι παράγοντες που επηρεάζουν το αποτέλεσμα του effect size που χρησιμοποιείται στην εκάστοτε μετα-ανάλυση. Για παράδειγμα έχουμε μια σειρά μελετών με δύο ομάδες πληθυσμού ανά μελέτη που ελέγχεται η υπόθεση αν η θεραπεία μέσω placebo φαρμάκων [2] είναι αποτελεσματικότερη από κάποιο εμβόλιο. Αν η θεραπεία που είναι υπό έλεγχο έχει αποτέλεσμα, το effect size που θα προκύψει θα είναι «παρόμοιο» σε όλες τις μελέτες που συμμετέχουν στην μετα-ανάλυση. Το effect size στην συγκεκριμένη μετα-ανάλυση μπορεί να είναι υψηλότερο ή χαμηλότερα επειδή επηρεάζεται από παράγοντες όπως η ηλικία, η σωματική υγεία του ασθενούς ή ακόμα και το φύλλο. Επειδή οι συμμετέχοντες στις μελέτες ποικίλουν όπως επίσης και τα χαρακτηριστικά αυτών θα έχουμε διαφορετικά effect sizes για διαφορετικές μελέτες.

Θα πρέπει να αναφερθεί ότι στις μελέτες πραγματοποιείται έλεγχος ετερογένειας και επιλέγοντας να χρησιμοποιηθεί random effect model αν δεν υπάρχει ετερογένεια είναι σαν να έχουμε fixed effect model δηλαδή οι πληθυσμοί είναι όμοιοι. Στην περίπτωση όμως που χρησιμοποιείται fixed effect model και από τον έλεγχο προκύψει ετερογένεια σημαίνει σημαίνει ότι θα προκύψουν εσφαλμένα αποτελέσματα. Είναι ένας σημαντικός έλεγχος για την διασφάλιση σωστών αποτελεσμάτων.

### 2.2 Κλασικές Μέθοδοι μετα-ανάλυσης

Η μέθοδος του Peto υπολογίζει το log odds ratio μέσα από πίνακες ενδεχομένων. Ο πίνακας αυτός περιέχει το σύνολο των ασθενών που παρουσιάζουν κάποιο γεγονός (θάνατος) και αυτών που δεν παρουσιάζουν κάποιο γεγονός, όπως φαίνεται στον πίνακα 2 παραγράφου 1.2.1. Ο χρόνος που χρειάζεται να ολοκληρωθεί η εκάστοτε μελέτη χωρίζεται σε μικρότερες χρονικές περιόδους και έτσι προκύπτουν διαφορετικοί πίνακες ενδεχομένων για κάθε μια από αυτές [21]. Το treatment effect υπολογίζεται σε αυτή την μέθοδο από το odds ratio.

Μια άλλη μέθοδος που χρησιμοποιεί τα odds ratio για τον υπολογισμό του treatment effect είναι αυτή που ονομάζεται μέθοδος log rank. Η συγκεκριμένη μέθοδος υλοποιείται όπως ακριβώς και του Peto, με την διαφορά ότι οι χρονικές περίοδοι είναι όσες το σύνολο των γεγονότων και κάθε γεγονός ορίζει το τέλος μιας χρονικής περιόδου [21]. Αυτό σημαίνει ότι όταν έχουμε γεγονός γράφεται 1 και όταν έχουμε αποχώρηση κάποιου από την μελέτη (censoring) γράφεται 0 στον εκάστοτε πίνακα ενδεχομένων. Η συγκεκριμένη μέθοδος λαμβάνει υπόψιν της τους censored ασθενείς καθώς με το που αποχωρήσουν από την μελέτη αφαιρούνται από τους υπολογισμούς των διαφόρων στατιστικών μεγεθών.

Επίσης υπάρχει και η μέθοδος υπολογισμού Interval-censored που χωρίζει την μελέτη σε χρονικές περιόδους όπου κάθε ένας που συμμετέχει στη μελέτη είτε επιβιώνει κατά την διάρκεια της χρονικής

περιόδου, είτε πεθαίνει, είτε αποχωρεί από την μελέτη. Τα δεδομένα μπορούν να θεωρηθούν σαν αποτελέσματα μελετών Bernoulli καθώς κάθε ασθενής έχει μια πιθανότητα να είναι σε κίνδυνο (at risk) κατά την διάρκεια του χρονικού διαστήματος που έχει οριστεί δηλαδή να εμφανίσει ή να μην εμφανίσει γεγονός. Για παράδειγμα αν κάποιος που συμμετέχει στη μελέτη και πεθαίνει (γεγονός) στον τρίτο χρόνο –με δεδομένο ότι η χρονική περίοδος που έχει οριστεί είναι το 1 έτος- έχει τρεις καταγραφές. Η πρώτη είναι τον πρώτο χρόνο που επέζησε, η δεύτερη καταγραφή είναι το δεύτερο χρόνο που ισχύει το ίδιο και τον τρίτο χρόνο είναι που παρουσίασε το γεγονός. Τα δεδομένα σε αυτή την μέθοδο αντιμετωπίζονται σαν interval censored δηλαδή είναι γνωστή η χρονική περίοδος που συνέβη ένα γεγονός αλλά όχι ο ακριβής χρόνος.

Στις μεθοδολογίες μετα-ανάλυσης που αναφέρθηκαν πριν ξεκινήσει η διαδικασία της μετα-ανάλυσης και εφόσον υπάρχουν τα IPD πραγματοποιείται ένας έλεγχος το λεγόμενο Log rank test. Το συγκεκριμένο test ελέγχει αν η συνάρτηση επιβίωσης που χρησιμοποιείται στους δύο πληθυσμούς είναι ίδια [17]. Σε περίπτωση που είναι ίδια η συγκεκριμένη μελέτη θα συμπεριληφθεί στην μετα-ανάλυση αλλιώς την απορρίπτουμε.

## 2.3 Εναλλακτικές μέθοδοι μετα-ανάλυσης

Μια εναλλακτική μέθοδος που χρησιμοποιεί δεδομένα από μετα-αναλύσεις με IPD με αποτελέσματα time-to-event είναι ο λόγος των ποσοστών (percentile ratio). Η συγκεκριμένη μέθοδος έχει προταθεί επειδή τα percentile ratios μπορεί να προσδιοριστούν και να χρησιμοποιηθούν για την σύγκριση οποιονδήποτε 2 καμπυλών επιβίωσης και το αποτέλεσμα που προκύπτει να αποτελεί την βάση για την σύγκριση και τον συνδυασμό των treatment effects για όλες τις μελέτες της μετα-ανάλυσης [22].

Υπάρχουν περιπτώσεις που δεν μπορεί να υπολογιστεί το HR σε κάποιες μελέτες ώστε να πραγματοποιηθεί η μετα-ανάλυση. Η αιτία που συμβαίνει αυτό είναι γιατί δεν υπάρχουν τα IPD ή γενικά είναι ανεπαρκή τα δεδομένα που υπάρχουν στην διάθεση του αναλυτή. Έτσι προτάθηκε η μέθοδος του μέσου λόγου (median ratio) που μπορεί να εφαρμοστεί με τρεις τρόπους. Η διακύμανση να υπολογιστεί κατά προσέγγιση χρησιμοποιώντας 1<sup>ov</sup> τους θανάτους στο control group και στο research group για κάθε μελέτη 2<sup>ov</sup> τους συνολικούς θανάτους σε κάθε μελέτη 3<sup>ov</sup> τον συνολικό αριθμό των ασθενών σε κάθε μελέτη [23]. Για να ελεγχθεί η αποτελεσματικότητα της μεθόδου τα αποτελέσματα από την μετα-ανάλυση συγκρίθηκαν έχοντας σαν effect size είτε το HR, είτε το MR, είτε το OR. Το τελικό συμπέρασμα της σύγκρισης της αρχικής μετα-ανάλυσης -που έγινε από IPD με effect size το HR- με τις άλλες 2 μεθόδους είναι ότι ο 1<sup>ος</sup> τρόπος του υπολογισμού της διακύμανσης μέσω του MR είναι αυτός είναι ο πιο αποτελεσματικός. Όταν όμως έχουμε στην διάθεση μας τα IPD πάντα επιλέγουμε το HR σαν effect size [23].

Οι αναλυτές γενικά προσπαθούν να αναπτύξουν μεθόδους που δεν χρειάζονται όλα τα δεδομένα (IPD) μιας μελέτης, γιατί συνήθως είναι δύσκολο ή και αδύνατο να εντοπιστούν. Έτσι έχουμε την ανάπτυξη μιας άλλης μεθόδου που χρησιμοποιεί στοιχεία που συνήθως αναφέρονται στις δημοσιευμένες μελέτες ώστε να υπάρχει η δυνατότητα να συμπεριληφθούν όσο τον δυνατόν περισσότερες σε μια μετα-ανάλυση. Η μεθοδολογία αυτή υπολογίζει ένα αθροιστικό ποσοστό επιβίωσης (συνάρτηση επιβίωσης) για τον κάθε πληθυσμό της εκάστοτε μελέτης. Συγκεκριμένα υπολογίζεται η διαφορά μεταξύ log(-log) αθροιστικού ποσοστού επιβίωσης για τις δύο θεραπείες που ελέγχονται στους δυο πληθυσμούς μιας μελέτης σαν δείκτης διαφοράς για να αναπαραστήσει το effect size της συγκεκριμένης μελέτης [24]

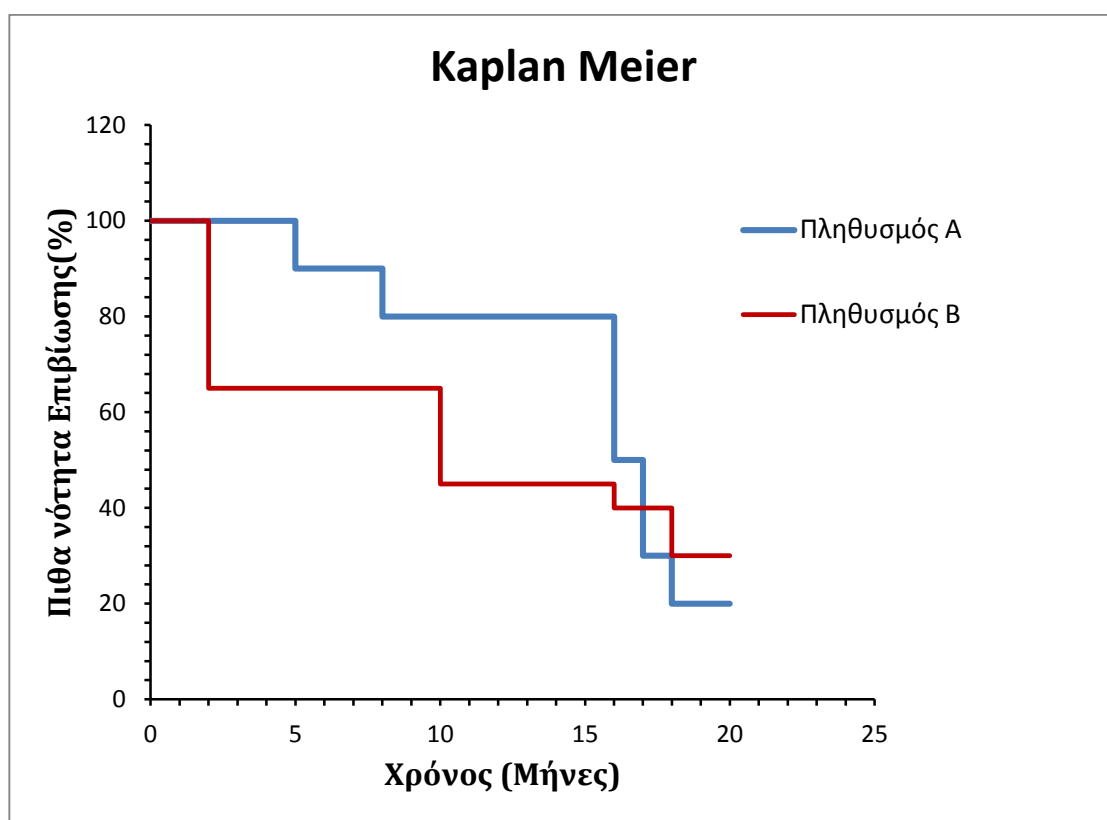
Μια ακόμη αξιοσημείωτη είναι αυτή που προτάθηκε από τον Preneger και είναι μια απλή διαδικασία που υπολογίζει το HR κατά προσέγγιση χρησιμοποιώντας τα σύνολα των γεγονότων και των μη γεγονότων από τους πίνακες ενδεχομένων [25]. Η μέθοδος είναι απλή αλλά έχει το μειονέκτημα ότι δεν λαμβάνονται υπόψη οι censored ασθενείς.

## 2.4 Μεθοδολογία μετα-ανάλυσης από δημοσιευμένες καμπύλες Kaplan Meier

Υπάρχουν αρκετοί τρόποι διεξαγωγής μετα-αναλύσεων αλλά στην συγκεκριμένη εργασία γίνεται μια προσπάθεια αξιοποίησης των εικόνων και των δεδομένων που έχουν δημοσιευτεί από διάφορες μελέτες που μπορεί να έχουν αρκετή ανεκμετάλλευτη πληροφορία. Έτσι στο συγκεκριμένο κεφάλαιο θα γίνει μια παρουσίαση της μεθοδολογίας της μετα-ανάλυσης μέσα από τα δεδομένα που δίνουν οι δημοσιευμένες καμπύλες επιβίωσης (Kaplan-Meier) καθώς επίσης και των μεταβλητών που διέπουν αυτή.

### 2.4.1 Καμπύλη επιβίωσης KAPLAN MEIER

Το 1958 έγινε μια σημαντική συνεργασία μεταξύ του Edward L. Kaplan και του Paul Meier για την δημοσίευση μιας εργασίας [20] στο περιοδικό the American Statistical Association με θέμα την αντιμετώπιση των ημιτελών παρατηρήσεων σε διάφορες μελέτες. Αυτή η συνεργασία οδήγησε στην ανάπτυξη μιας πολύ σημαντικής μεθόδου αντιμετώπισης μη παραμετρικών υπολογισμών και ημιτελών παρατηρήσεων που βασίστηκε στην θεωρία της πιθανότητας υπό συνθήκη. Πρόκειται για μια μέθοδο που υπολογίζει την πιθανότητα επιβίωσης σε μια συγκεκριμένη χρονική στιγμή χρησιμοποιώντας το ποσοστό των ασθενών που έχουν επιβιώσει μέχρι και εκείνη την στιγμή [21]. Ο υπολογισμός της πιθανότητας επιβίωσης πραγματοποιείται λαμβάνοντας υπόψη τα γεγονότα που είναι ουσιαστικά ο θάνατος κάποιου ασθενή κατά την διάρκεια της μελέτης και τα «censored» δεδομένα της μελέτης που ουσιαστικά είναι ασθενείς που είτε αποχωρούν για κάποιο λόγο από την μελέτη είτε σταματούν να εμφανίζονται και δεν υπάρχει κανένα στοιχείο για να βρεθούν. Συνήθως χρησιμοποιείται για την σύγκριση δυο διαφορετικών πληθυσμών και την μεταξύ τους σχέση όπως φαίνεται και στην εικόνα 3. Η καμπύλη που προκύπτει από την εφαρμογή της μεθόδου Kaplan-Meier και ο τρόπος αναπαράστασης της είναι ένας συνδυασμός όλων όσων αναφέρθηκαν και θα παρουσιαστούν με περισσότερες λεπτομέρειες στη συνέχεια.



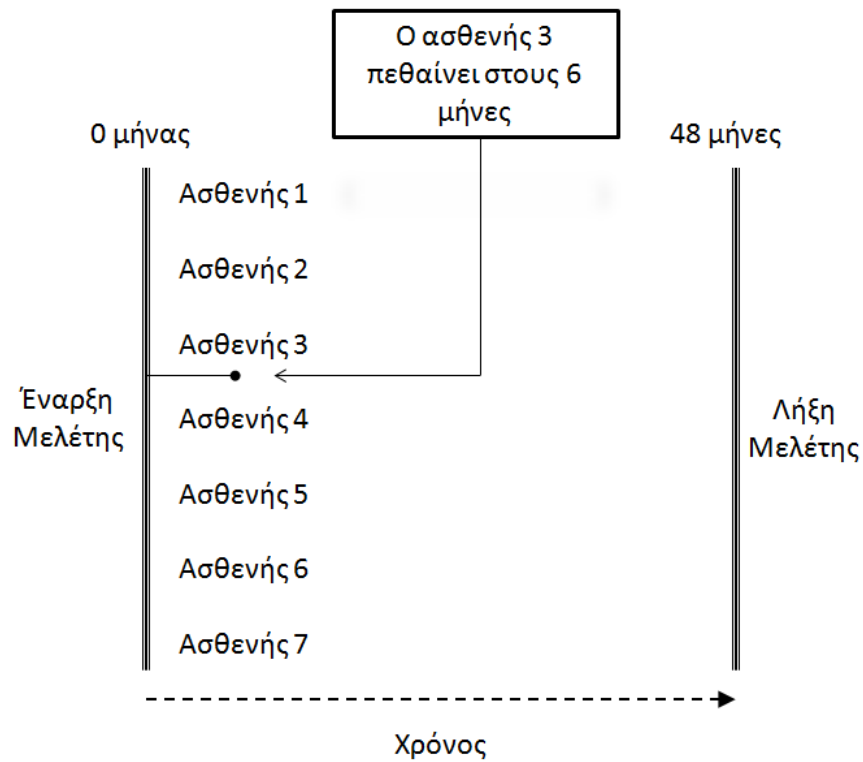
Εικόνα 3 Είναι ένα βασικό μοντέλο απεικόνισης της πιθανότητας επιβίωσης που βρέθηκε με την μέθοδο Kaplan Meier.

#### 2.4.1.1 Δημιουργία καμπύλης

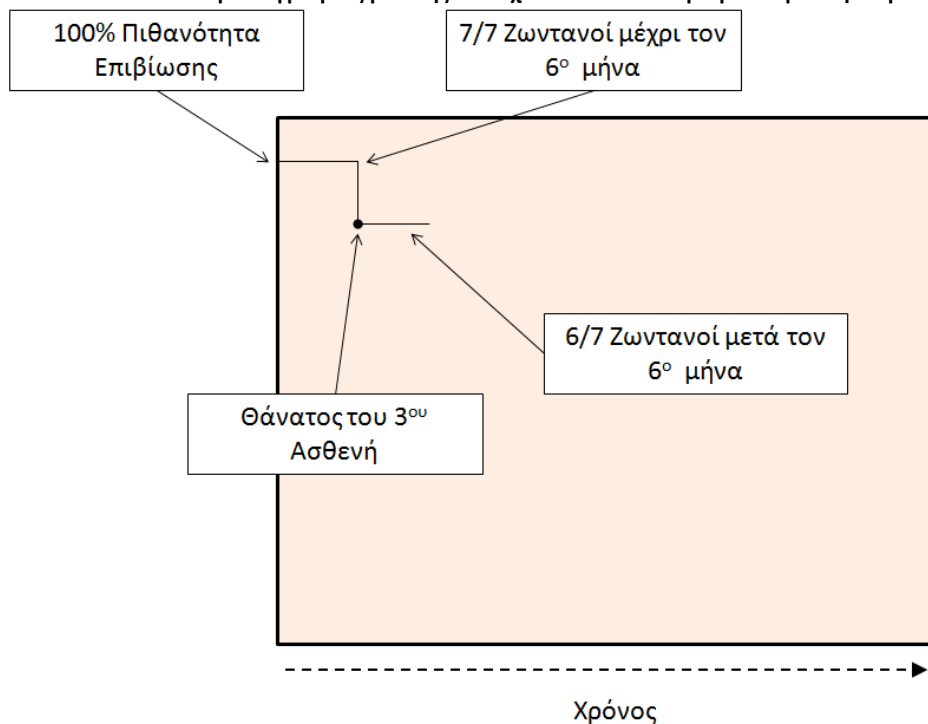
Το παράδειγμα που ακολουθεί θα διευκολύνει την κατανόηση για τον τρόπο που σχηματίζεται η καμπύλη επιβίωσης Kaplan Meier [22]. Έστω 7 ασθενείς που συμμετέχουν σε μια μελέτη που διαρκεί 48 μήνες. Ακολουθεί μια συνοπτική παρουσίαση των σταδίων δημιουργίας καμπυλών Kaplan-Meier.

- Στην μελέτη θα συμμετάσχει ένας συγκεκριμένος αριθμός ατόμων με αρχική πιθανότητα επιβίωσης ίση 100%
- Με την εμφάνιση του πρώτου θανάτου θα μειωθεί ο αριθμός των ασθενών και θα αλλάξει και η πιθανότητα επιβίωσης
- Στον υπολογισμό της πιθανότητας επιβίωσης εκτός από τα γεγονότα που είναι οι θάνατοι σημαντικός ρόλο παίζουν και οι «censored» ασθενείς
- Με βάση την εμφάνιση των γεγονότων και των censored ασθενών σχηματίζεται μια φθίνουσα καμπύλη KM

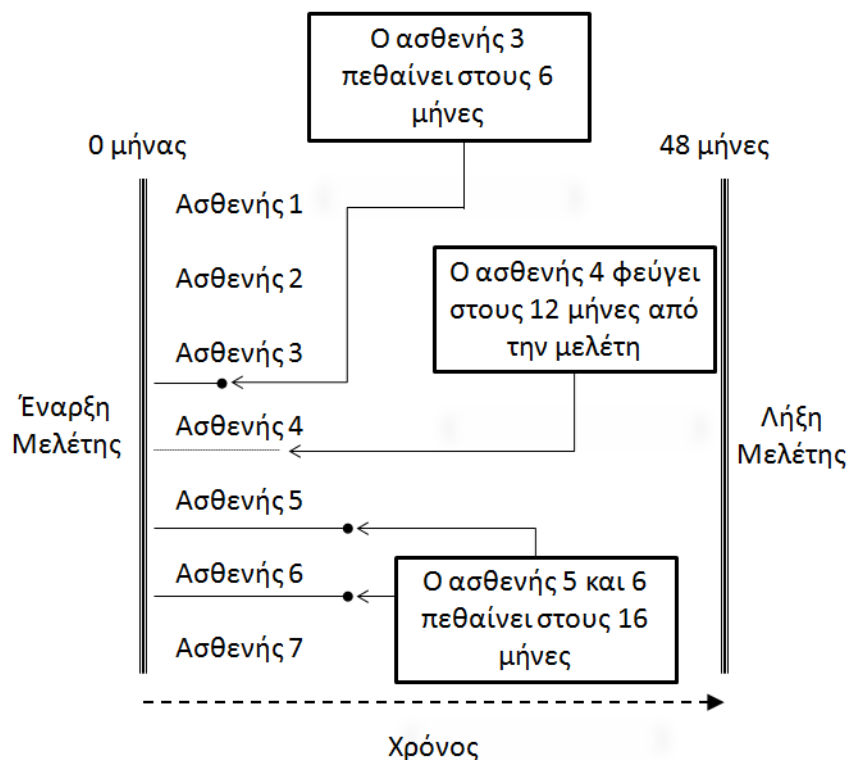




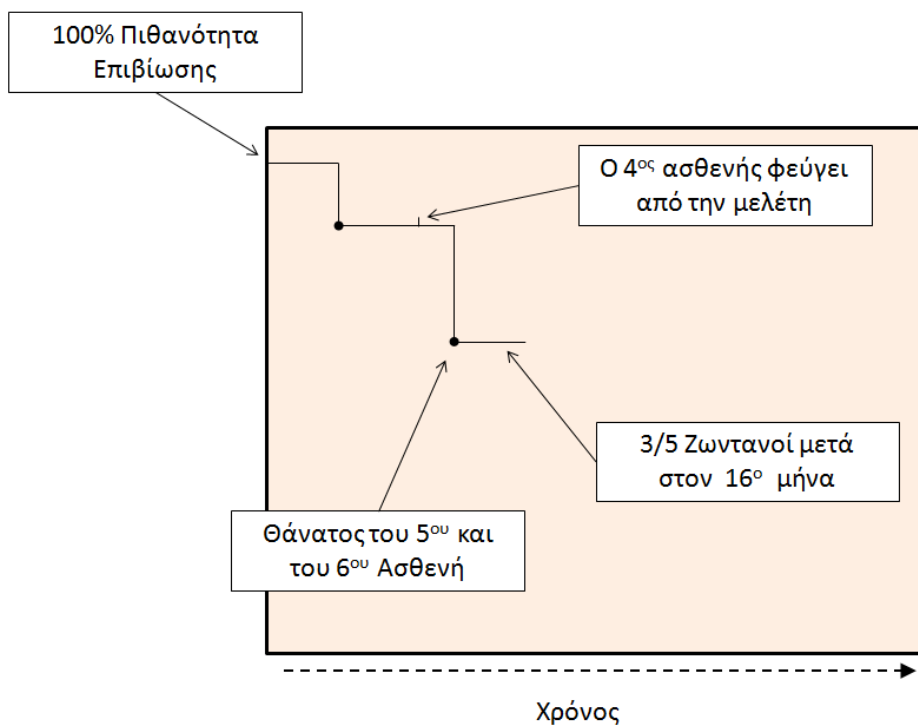
Εικόνα 4 Παράδειγμα μιας μελέτης που έχει σαν αποτέλεσμα μια καμπύλη επιβίωσης KM.



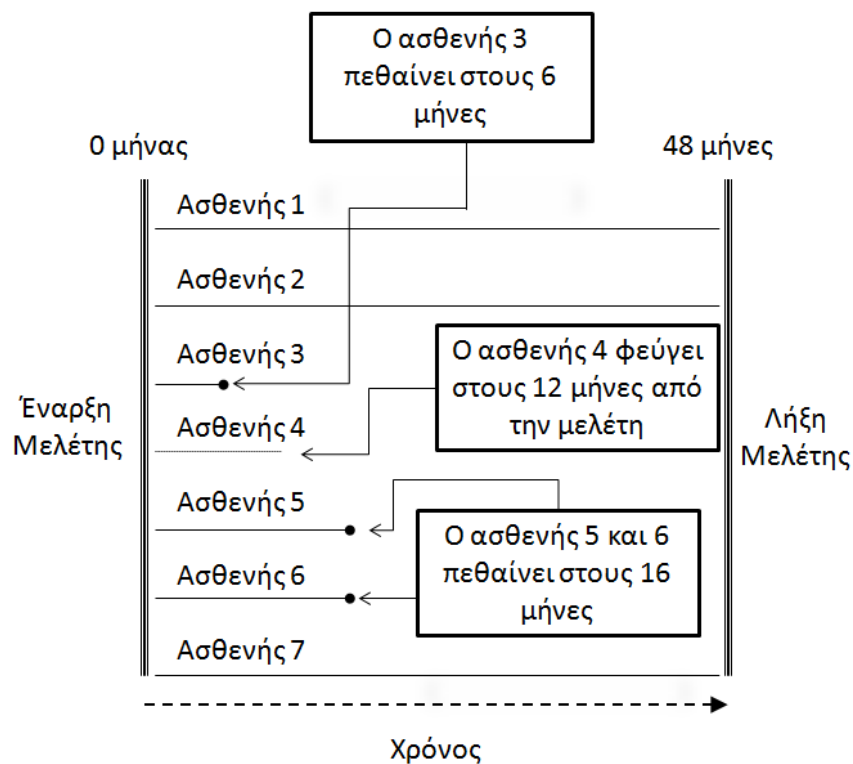
Εικόνα 5 Διαμόρφωση της καμπύλης μετά το πρώτο συμβάν στην μελέτη.



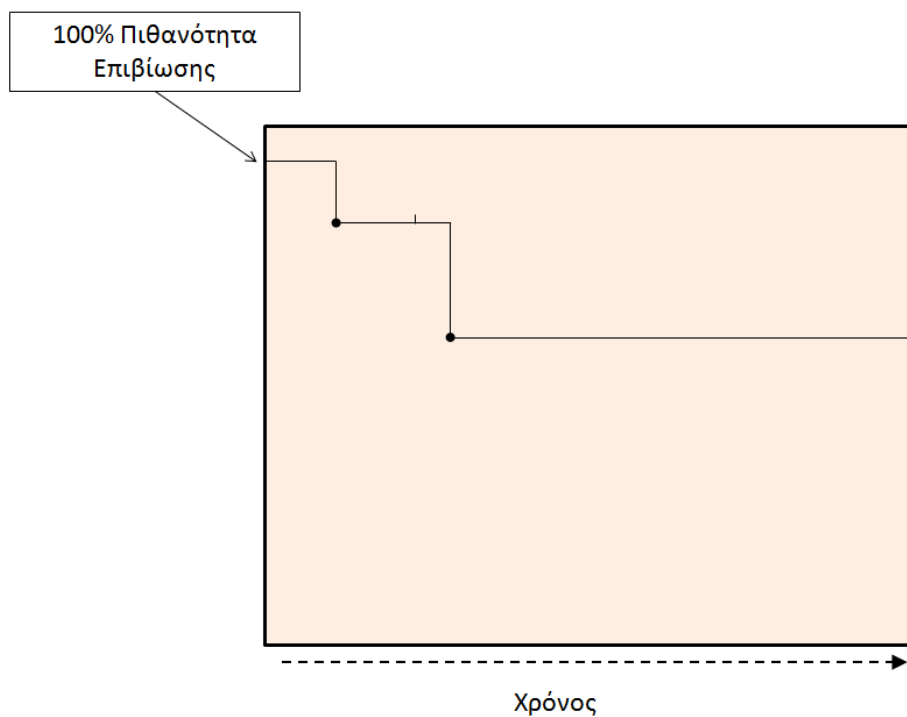
Εικόνα 6 Εμφάνιση και άλλων γεγονότων και αποχώρηση κάποιου ασθενή από την μελέτη.



Εικόνα 7 Η μορφή της καμπύλης επιβίωσης μέχρι και τον 16<sup>ο</sup> μήνα πραγματοποίησης της μελέτης.



Εικόνα 8 Ολοκλήρωση της μελέτης μετά το πέρας των 48 μηνών.



Εικόνα 9 Τελικό αποτέλεσμα καμπύλης επιβίωσης με την μέθοδο K-M.

#### 2.4.1.2 Μεταβλητές της μεθόδου K-M

Έστω ότι υπάρχουν  $N$  ασθενείς οι οποίοι συμμετέχουν σε μια μελέτη που διαρκεί  $T$  χρόνο. Ο συνολικός χρόνος πραγματοποίησης της μελέτης χωρίζεται σε μικρότερα χρονικά σημεία  $\tau_j$  με το  $j = 1, 2, \dots, k$  και το  $k$  το σύνολο των χρονικών σημείων τα οποία ορίζουν την αρχή και την λήξη των χρονικών διαστημάτων  $t$  (πχ. χρονικό διάστημα  $(\tau_j, \tau_{j+1})$ ). Το πλήθος των διαστημάτων και το μήκος κάθε χρονικού διαστήματος μπορεί να διαφέρει επειδή διαμορφώνεται ανάλογα με το ρυθμό εμφάνισης των γεγονότων που συμβαίνουν κατά την διάρκεια της μελέτης. Στην πράξη τα χρονικά διαστήματα ορίζονται έτσι ώστε κάθε ένα από αυτά να περιλαμβάνει μικρό αριθμό γεγονότων. Συμβολίζουμε με  $D_j$  το πλήθος των γεγονότων τα οποία λαμβάνουν χώρα κατά την διάρκεια του χρονικού διαστήματος  $(\tau_j, \tau_{j+1})$ . Οι ασθενείς που βρίσκονται σε κίνδυνο (at risk) δηλαδή τα άτομα του πληθυσμού που είναι ζωντανά και συμμετέχουν στην μελέτη συμβολίζονται με το  $R_j$  και υπολογίζονται κατά την διάρκεια του χρονικού διαστήματος  $j$ . Οι censored [21] ασθενείς παίζουν επίσης σημαντικό ρόλο στους παρακάτω υπολογισμούς, συμβολίζονται με  $C_j$  και λαμβάνονται υπόψιν για το χρονικό διάστημα  $(\tau_j, \tau_{j+1})$ . Ο υπολογισμός της πιθανότητας επιβίωσης χωρίς censoring προκύπτει από την σχέση 13.

$$\hat{S}(t) = \frac{\text{αριθμός ασθενών που έχουν επιβιώσει μετά το } t}{\text{συνολικός αριθμός των ασθενών}} \quad (13)$$

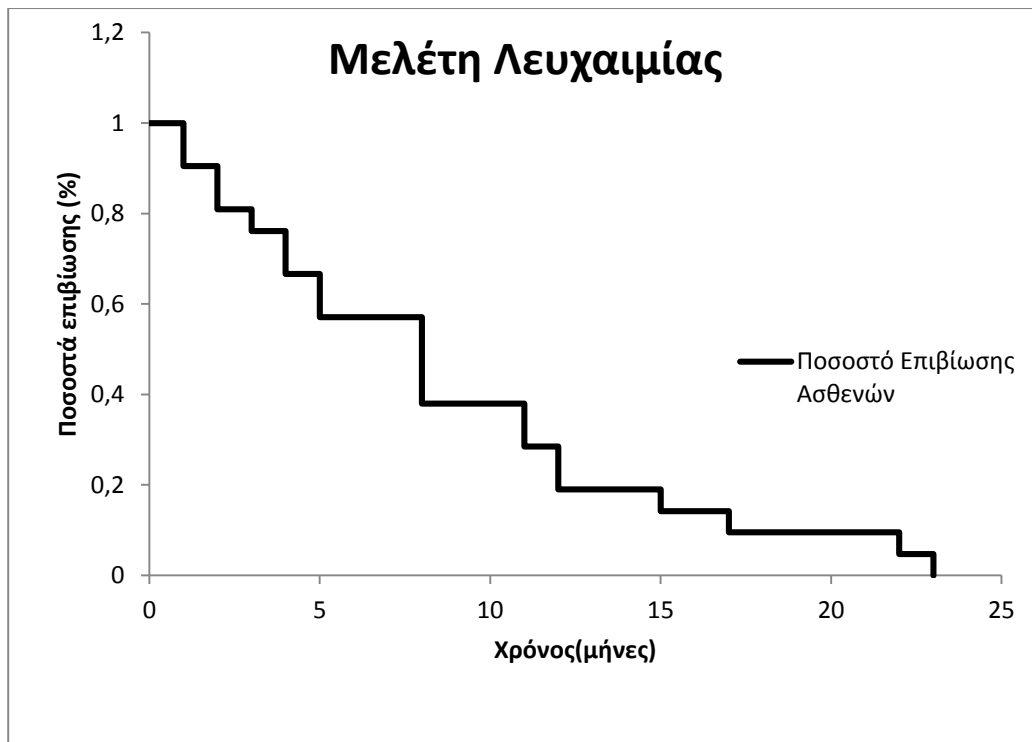
#### 2.4.1.3 Παράδειγμα εφαρμογής K-M χωρίς να υπάρχει censoring

Έστω 21 ασθενείς που δέχονται μια συγκεκριμένη [23] θεραπεία για την λευχαιμία με τον πίνακα 3 που περιέχει τα στατιστικά στοιχεία της μελέτης και την εικόνα 10 τα αποτελέσματα σε καμπύλη K-M. Περιγραφή του πίνακα:

- Η πρώτη στήλη περιέχει το μήκος των χρονικών διαστημάτων σε μήνες που έχουν διαμορφωθεί από τις χρονικές στιγμές των θανάτων των ασθενών ανάλογα με τον σχεδιασμό της μελέτης.
- Η δεύτερη στήλη περιέχει τους ζωντανούς ασθενείς κατά την έναρξη του χρονικού διαστήματος που αντιστοιχεί.
- Η Τρίτη στήλη περιέχει το σύνολο των ασθενών που πεθαίνουν κατά την διάρκεια της μελέτης στα συγκεκριμένα χρονικά διαστήματα
- Η τέταρτη στήλη περιέχει της πιθανότητες επιβίωσης που αντιστοιχούν σε συγκεκριμένα χρονικά διαστήματα.

Πίνακας 3 Αποτελέσματα μελέτης σε ασθενείς με λευχαιμία που διήρκεσε 23 μήνες και κανένας από τους ασθενείς δεν έφυγε από την μελέτη (δλδ. δεν υπήρχε censoring) .

Χρονικά Διαστήματα (σε μήνες)	Ζωντανοί Ασθενείς στην Αρχή των Χρονικών Διαστημάτων	Ασθενείς που Πεθαίνουν Κατά την Διάρκεια των Διαστημάτων	Ποσοστό επιβίωσης ανά χρονικό διάστημα
$t$	$R$	$D$	$\hat{S}(t)$
0-1	21	2	1,000
1-2	19	2	0,905
2-3	17	1	0,809
3-4	16	2	0,761
4-5	14	2	0,666
5-8	12	4	0,571
8-11	8	2	0,380
11-12	6	2	0,285
12-15	4	1	0,190
15-17	3	1	0,142
17-22	2	1	0,095
22-23	1	1	0,047



Εικόνα 10 Αποτελέσματα μελέτης λευχαιμίας σε καμπύλη επιβίωσης Kaplan Meier.

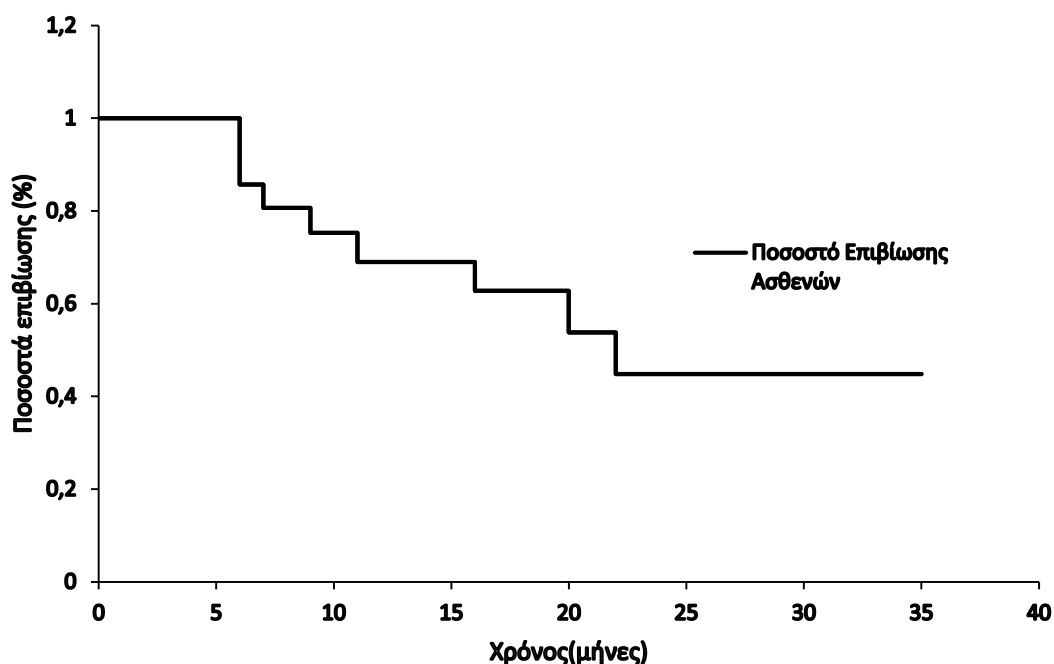
#### 2.4.1.4 Παράδειγμα εφαρμογής K-M με censoring

Ο υπολογισμός των πιθανοτήτων επιβίωσης που στην περίπτωση που έχουμε censoring χρησιμοποιούμε την σχέση 14.

$$\hat{S}(t) = \prod_{j: \tau_j < t} \left(1 - \frac{d_j}{r_j}\right) \quad (14)$$

Ο τύπος αυτός εφαρμόζεται για τα  $k - 1$  χρονικά διαστήματα εφόσον τα χρονικά σημεία που τα σχηματίζουν είναι  $k$ . Τα χρονικά διαστήματα έχουν οριστεί όπως αναφέρθηκε προηγουμένως από το χρονικό σημείο έναρξης μέχρι και την ολοκλήρωση της μελέτης. Τα δεδομένα που δίνονται [23] είναι +6, 6, 6, 6, 7, +9, +10, 10, +11, 13, 16, +17, +19, +20, 22, 23, +25, +32, +32, +34, +35 με το + να συμβολίζει τους μήνες που παρουσιάζονται censored ασθενείς και τα υπόλοιπα να είναι οι μήνες που παρουσιάζετε κάποιος θάνατος (γεγονός). Από τα δεδομένα αυτά προκύπτει ο πίνακας 4 και η εικόνα 11. Είναι προφανές ότι οι censored ασθενείς δεν συμμετέχουν στους υπολογισμούς αλλά αυτό που επηρεάζει το αποτέλεσμα είναι οι ασθενείς σε κίνδυνο στον χρόνο που έχει παρέλθει.

## Μελέτη Λευχαιμίας



Εικόνα 11 Η καμπύλη επιβίωσης που προκύπτει με βάση τα στοιχεία του πίνακα 2 .

Πίνακας 4 Αποτελέσματα από την μελέτη που πραγματοποιήθηκε σε άτομα που έπασχαν από λευχαιμία.

Χρονικά διαστήματα $t$	Θάνατοι ανά χρονικό διάστημα $D$	Ασθενείς censored ανά χρονικό διάστημα $C$	Ασθενείς σε κίνδυνο ανά χρονικό διάστημα $R$	Ποσοστό επιβίωσης ανά χρονικό διάστημα $S$
0-6	3	1	21	0,857143
6-7	1	0	17	0,806723
7-9	0	1	16	0,806723
9-10	1	1	15	0,752941
10-11	0	1	13	0,752941
11-13	1	0	12	0,690196
13-16	1	0	11	0,627451
16-17	0	1	10	0,627451
17-19	0	1	9	0,627451
19-20	0	1	8	0,627451
20-22	1	0	7	0,537815
22-23	1	0	6	0,448179
23-25	0	1	5	0,448179
25-32	0	1	4	0,448179
32-34	0	1	2	0,448179
34-35	0	1	1	0,448179

## 2.5 Μέθοδος μετα-ανάλυσης από δημοσιευμένες καμπύλες

Έπειτα από ενδελεχή έρευνα στη βιβλιογραφία και σε ιστοσελίδες με επιστημονικά περιοδικά και μελέτες βρέθηκαν αρκετά στοιχεία και δεδομένα που έχουν σαν αντικείμενο την διαδικασία της μετα-ανάλυσης και τις καμπύλες Kaplan Meier. Η αξιοποίηση των στοιχείων αυτών οδηγεί σε χρήσιμα συμπεράσματα. Για να επιτευχθεί αυτός ο στόχος επιλέχθηκε η μεθοδολογία που θα παρουσιαστεί στην παρούσα εργασία και μας δίνει την δυνατότητα να αντλούμε δεδομένα και στοιχεία από αυτές τις καμπύλες ώστε να τα χρησιμοποιήσουμε για την πραγματοποίηση μετα-ανάλυσης.

- I. Θα γίνει μια παρουσίαση της γενικής προσέγγισης για να επιτευχθεί η μετα-ανάλυση με την βοήθεια δημοσιευμένων καμπυλών. Πριν γίνει η παρουσίαση της μεθόδου είναι πολύ σημαντικό να δοθεί έμφαση σε κάποιες βασικές λεπτομέρειες. Για κάθε μια από τις μελέτες που θα χρησιμοποιηθούν για την μετα-ανάλυση θα οριστούν διαφορετικά χρονικά διαστήματα που όλα μαζί θα αποτελούν ουσιαστικά τον χρόνο  $T$  που χρειάστηκε για να πραγματοποιηθεί η μελέτη και θα είναι συνεχόμενα και μη επικαλυπτόμενα. Έπειτα θα υπολογιστεί το Log Hazard Ratio για κάθε ένα από αυτά τα διαστήματα και συνδυάζοντας τα με συγκεκριμένο τρόπο θα εξαχθεί ένα συνολικό αποτέλεσμα για κάθε μελέτη. Τα βήματα της διαδικασίας παρουσιάζονται αναλυτικά παρακάτω. Αρχικά πραγματοποιείται ο υπολογισμός των πιθανοτήτων επιβίωσης για την μελέτη  $i$  ( $i = 1, \dots, K$ ) μετά την επεξεργασία της εικόνας που περιέχει την αντίστοιχη καμπύλη επιβίωσης Kaplan-Meier. Η διαδικασία εύρεσης των πιθανοτήτων επιβίωσης περιγράφεται με λεπτομέρειες στο υποκεφάλαιο αυτής της ενότητας. Εφόσον εξετάζουμε δύο πληθυσμούς control group και research group που αντιστοιχούν στην εικόνα στις δυο διαφορετικές καμπύλες, υπολογίζονται ξεχωριστά οι πιθανότητες επιβίωσης για κάθε μια από αυτές. Ο υπολογισμός των πιθανοτήτων επιβίωσης [16] πραγματοποιείται για κάθε ένα από τα ορισμένα χρονικά διαστήματα ( $t=1, \dots, T$ ) που ουσιαστικά είναι τα σημεία έναρξης και λήξης των χρονικών διαστημάτων. Στο κείμενο που θα συνοδεύει την καμπύλη υπάρχουν στοιχεία που θα μας βοηθήσουν να υπολογίσουμε την μέγιστη και την ελάχιστη περίοδο συμμετοχής των ασθενών στην εκάστοτε μελέτη καθώς όπως προαναφέρθηκε είναι ένας πολύ βασικός ο λόγος που χρησιμοποιούνται η μεθοδολογία KM. Πολύ σημαντικό για την κατανόηση της μεθόδου και των επόμενων βημάτων είναι ο σαφής διαχωρισμός των ασθενών σε κίνδυνο στην αρχή του χρονικού διαστήματος με τους ασθενείς σε κίνδυνο κατά την διάρκεια του. Με τους τελευταίους να επιτρέπει το censoring κατά την διάρκεια των χρονικών διαστημάτων. Για να εξαλειφθούν όλες οι πιθανότητες σύγχυσης θα χρησιμοποιηθούν επιπλέον βοηθητικά στοιχεία. Συγκεκριμένα για το χρονικό διάστημα  $(t - 1, t)$  χρησιμοποιούμε τον συμβολισμό  $t_s$  και  $t_e$  για να αναπαραστήσουμε την αρχή και το τέλος του διαστήματος και με το  $t$  να συμβολίζει όλο το χρονικό διάστημα  $(t - 1, t)$ .
- II. Υπολογίζουμε τους ασθενείς που είναι ζωντανοί δηλαδή αυτούς που είναι σε κίνδυνο για κάθε χρονικό διάστημα. Αυτό θα επιτευχθεί χρησιμοποιώντας ένα μοντέλο υπολογισμού των censored ασθενών κατά την διάρκεια ενός χρονικού διαστήματος. Το μοντέλο αυτό χρησιμοποιεί την μέγιστη και την ελάχιστη περίοδο συμμετοχής-παρακολούθησης (max και min follow up αντίστοιχα) των ασθενών στη μελέτη για να υπάρχει η δυνατότητα υπολογισμού των στατιστικών μεγεθών που μας ενδιαφέρουν. Αν υποθέσουμε ότι ο ρυθμός των censored ασθενών σε μια μελέτη είναι σταθερός και δεν περιέχει πληροφορίες, τότε δεδομένων των τυχαίων αριθμών και της διάρκειας των παρακολουθήσεων μπορεί να υπολογιστεί ο αριθμός των ασθενών σε κίνδυνο.

Αυτό στην πραγματικότητα λειτουργεί σαν τον αριθμό των ασθενών με πλήρη παρακολούθηση που είναι ισοδύναμος προς το άθροισμα των μερικών παρακολουθήσεων για ένα συγκεκριμένο διάστημα [16]. Για παράδειγμα αν είχαμε 100 ασθενείς που σταματούν να συμμετέχουν στη μελέτη (censored) με ένα σταθερό ρυθμό για ένα χρόνο, τότε ο μέσος όρος παρακολούθησης είναι έξι μήνες και ο αριθμός ασθενών που συμμετέχουν στην μελέτη είναι 50, που ισούται με τους μισούς από όλους όσους ξεκίνησαν να συμμετέχουν στην μελέτη. Έτσι προκύπτει ότι όταν το μέγιστο και το ελάχιστο διάστημα παρακολούθησης είναι γνωστά και υποθέσουμε ότι μετά το μικρότερο διάστημα παρακολούθησης οι censored ασθενείς εμφανίζονται με ένα σταθερό ρυθμό μέχρι το μεγαλύτερο διάστημα παρακολούθησης. Ο αριθμός των ασθενών που παρακολούθησαν όλη την μελέτη μέχρι την δεδομένη χρονική στιγμή μπορούν να αναπαρασταθούν όπως φαίνεται στην εικόνα 11.

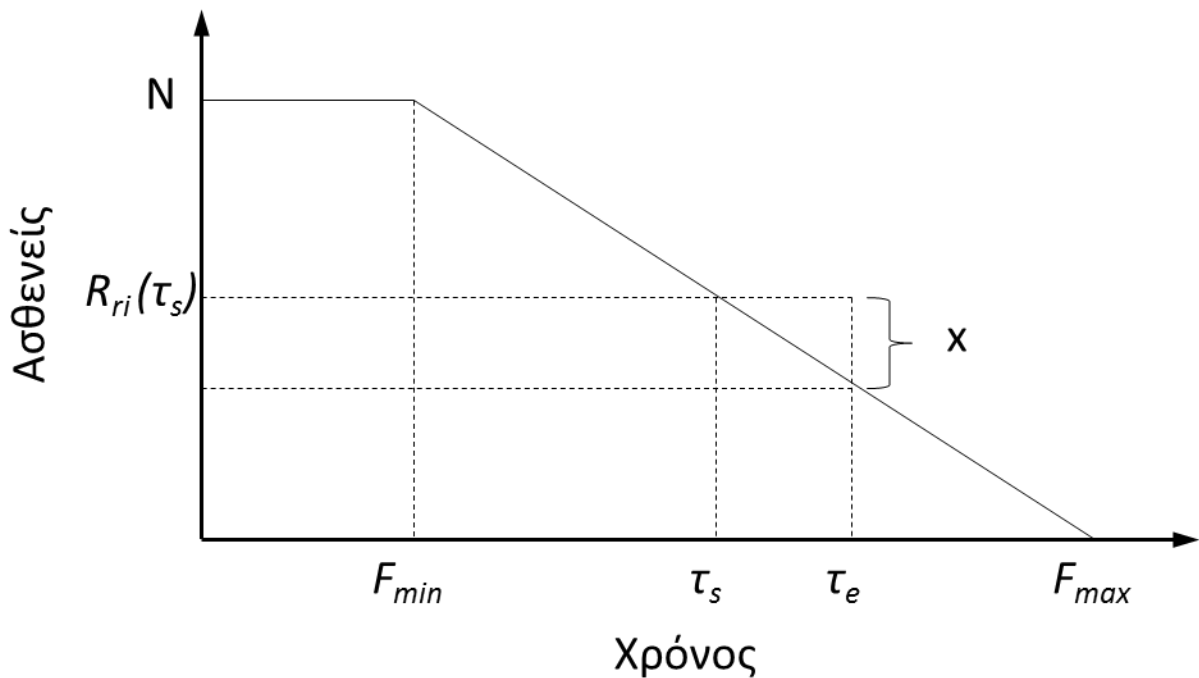
Παρακάτω παρουσιάζεται ένας αλγόριθμος για τον υπολογισμό του research group (ο ίδιο ισχύει και για το control group) για το χρονικό διάστημα  $(t-1, t)$  με το  $N$  να είναι ο συνολικός αριθμός των ασθενών που ανήκουν στον παραπάνω πληθυσμό, το  $F_{min}$  είναι το μικρότερο διάστημα παρακολούθησης στη μελέτη, το  $F_{max}$  είναι το μεγαλύτερο διάστημα παρακολούθησης στη μελέτη,  $t_s$  είναι το σημείο έναρξης του χρονικού διαστήματος  $(t-1, t)$ ,  $t_e$  είναι το σημείο λήξης του χρονικού διαστήματος  $(t-1, t)$  και το  $R_{ri}(t_s)$  είναι ο αριθμός των ασθενών σε κίνδυνο από το research group στην αρχή του χρονικού διαστήματος  $(t-1, t)$ .

Αν  $t_s \geq F_{min}$  και  $F_{min} \leq t_e \leq F_{max}$  τότε προκύπτει ότι ο αριθμός των ασθενών που είναι censored κατά την διάρκεια του διαστήματος  $(t-1, t) = \frac{1}{2}x$  (υποθέτοντας ότι οι censored ασθενείς εμφανίζονται με σταθερό ρυθμό κατά την διάρκεια ενός χρονικού διαστήματος) υπολογίζεται από

$$\begin{aligned} \frac{(F_{max} - t_s)}{R_{ri}(t_s)} &= \frac{(t_e - t_s)}{x} \\ \Rightarrow x &= R_{ri}(t_s) \frac{(t_e - t_s)}{(F_{max} - t_s)} \end{aligned}$$

Τελικά ο αριθμός των censored ασθενών που κατά την διάρκεια του διαστήματος  $(\tau_{j-1}, \tau_j)$  είναι

$$= R_{ri}(t_s) \left\{ \frac{1}{2} \frac{(t_e - t_s)}{(F_{max} - t_s)} \right\} \quad (15)$$



Εικόνα 12 Καμπύλη που αναπαριστά το πλήθος των συμμετεχόντων σε συνάρτηση με το χρόνο.

Αυτή η μέθοδος όπως προαναφέρθηκε υιοθετεί το γεγονός ότι οι censored ασθενείς έχουν σταθερό ρυθμό εμφάνισης κατά την διάρκεια ενός χρονικού διαστήματος και έτσι έχουν σταθερή συνεισφορά στον υπολογισμό του κινδύνου. Με λίγα λόγια ο αριθμός των ασθενών που κινδυνεύουν να πεθάνουν στο research group κατά την διάρκεια του χρονικού διαστήματος  $(t-1, t)$  δίνεται από την σχέση 16.

$$R_{ri}(t_s) = R_{ri}(t-1) - D_{ri}(t-1) - C_{ri}(t) \quad (16)$$



Όπου το  $R_{ri}(t-1)$  είναι ο αριθμός των ασθενών που βρίσκονται σε κίνδυνο στον πληθυσμό που είναι υπό έρευνα (research group) κατά την διάρκεια του χρονικού διαστήματος  $(t-2, t-1)$ . Το  $D_{ri}(t-1)$  είναι ο αριθμός των θανάτων στον πληθυσμό υπό έρευνα κατά την διάρκεια του χρονικού διαστήματος  $(t-2, t-1)$ , και το  $C_{ri}(t)$  είναι ο αριθμός των censored ασθενών από τον πληθυσμό υπό έρευνα κατά την διάρκεια του χρονικού διαστήματος  $(t-2, t-1)$  με τις παραπάνω σχέσεις και ορισμούς να ισχύουν και για τον control group. Για λόγου διευκόλυνσης δεχόμαστε ότι το  $R_{ri}(0)$  και το  $R_{ci}(0)$  είναι ο συνολικός αριθμός των ασθενών που συμμετέχουν στην μελέτη για είτε στον πληθυσμό υπό έρευνα είτε στον πληθυσμό ελέγχου (control group). Επιπλέον ότι για τα μεγέθη ισχύει  $D_{ri}(0) = D_{ci}(0) = C_{ri}(0) = C_{ci}(0) = 0$ .

Για να υπολογισθεί ο αριθμός των censored ασθενών κατά την διάρκεια κάποιου χρονικού διαστήματος  $(t-1, t)$  είναι απαραίτητο να γνωρίσουμε τον αριθμό των ασθενών που είναι ζωντανοί και σε κίνδυνο στην αρχή του συγκεκριμένου διαστήματος. Για τον πληθυσμό υπό έρευνα (research group) το μέγεθος αυτό προκύπτει από τη σχέση 17 η οποία ισχύει και για τον πληθυσμό ελέγχου (control group).

$$R_{ri}(t_s) = R_{ri}(t-1) - D_{ri}(t-1) \quad (17)$$

Ο αριθμός των censored ασθενών κατά την διάρκεια του χρονικού διαστήματος  $(t-1, t)$  υπολογίζεται με τον αλγόριθμο που περιγράφηκε παραπάνω στον οποίο χρησιμοποιήθηκαν τα μεγέθη μέγιστης και ελάχιστης παρακολούθησης. Από την στιγμή που έχει υπολογιστεί το σύνολο των ασθενών σε κίνδυνο του χρονικού διαστήματος  $(t-1, t)$  και ο αριθμός των censored ασθενών κατά την διάρκεια του ίδιου διαστήματος έχουμε την δυνατότητα να υπολογίσουμε το σύνολο των ασθενών σε κίνδυνο κατά την διάρκεια του διαστήματος αυτού. Για τον πληθυσμό υπό έρευνα αυτό προκύπτει από την σχέση 18 που ισχύει και για τον πληθυσμό ελέγχου.

$$R_{ri}(t) = R_{ri}(t_s) - C_{ri}(t) \quad (18)$$

Θα πρέπει να σημειωθεί ότι αν οι αριθμοί των ασθενών σε κίνδυνο δημοσιεύονται μαζί με τις καμπύλες επιβίωσης τότε αυτά μπορούν να χρησιμοποιηθούν ως εκτιμήσεις των  $R_{ri}(t_s)$  και  $R_{ci}(t_s)$  στους κατάλληλους χρόνους και μπορεί επίσης να χρησιμεύσουν σαν οδηγός για έχουμε ακριβέστερα αποτελέσματα βάση των μεθόδων που προτείνονται. Ο αριθμός των θανάτων στο χρονικό διάστημα  $(t-1, t)$  για τον πληθυσμό υπό έλεγχο μπορεί να υπολογισθεί από τη σχέση 19

$$D_{ri} = \left[ R_{ri}(t) \frac{S_{ri}(t_s) - S_{ri}(t_e)}{S_{ri}(t_s)} \right] \quad (19)$$

- III. Όπου το  $S_{ri}(t_s)$  είναι η εκτίμηση της πιθανότητας επιβίωσης του πληθυσμού υπό έλεγχο στην αρχή του χρονικού διαστήματος  $(t-1, t)$  που έχει υπολογισθεί με βάση τις δημοσιευμένες καμπύλες Kaplan Meier και  $S_{ri}(t_s)$  είναι η εκτίμηση της πιθανότητας επιβίωσης του πληθυσμού υπό έλεγχο στο τέλος του χρονικού διαστήματος  $(t-1, t)$  στην ίδια καμπύλη. Σύμφωνα με τα παραπάνω και με βάση την συγκεκριμένη μεθοδολογία που ακολουθείται όταν έχουμε censored ασθενείς και υπολογισμό χρόνου μέχρι το γεγονός (time to event) κατά την διάρκεια μιας μελέτης μπορούμε να υπολογίσουμε το Hazard Ratio από τα αποτελέσματα του σχετικού κινδύνου (relative risk). Έτσι προκύπτει ο υπολογισμός του log Hazard Ratio κατά την διάρκεια του χρονικού διαστήματος  $(t-1, t)$  από τη σχέση 20 και της από την σχέση 21.

$$\ln(HR_i(t)) = \ln\left(\frac{D_{ri}(t)/R_{ri}(t)}{D_{ci}(t)/R_{ci}(t)}\right) \quad (20)$$

$$\text{var}[\ln(HR_i(t))] = \frac{1}{D_{ri}(t)} - \frac{1}{R_{ri}(t)} + \frac{1}{D_{ci}(t)} - \frac{1}{R_{ci}(t)} \quad (21)$$

Υπάρχουν περιπτώσεις στις οποίες ο υπολογισμός του Log Hazard Ratio και της διακύμανσης δεν θα είναι εφικτός επειδή το  $D_{ci}(t)$  ή  $D_{ri}(t)$  μπορεί να είναι ίσο με το μηδέν και αυτό συμβαίνει όταν δεν έχουμε κανένα θάνατο στο χρονικό διάστημα που πραγματοποιούνται οι υπολογισμοί. Κάτω από αυτές τις συνθήκες για να πραγματοποιηθεί ο υπολογισμός στο συγκεκριμένο χρονικό διάστημα αντικαθιστούμε το 0 με κάποιο πολύ μικρό αριθμό και συγκεκριμένα με το  $10^{-6}$ . Αυτός ο αριθμός δεν είναι τυχαίος είναι αποδεδειγμένο ότι σε τέτοιες περιπτώσεις αν χρησιμοποιηθεί προκύπτει το καλύτερο δυνατό αποτέλεσμα. Η συγκεκριμένη προσέγγιση είναι προτιμότερη και από την συνένωση των διαστημάτων που θα μπορούσε να χρησιμοποιηθεί για να μην έχουμε 0 θανάτους.

- IV. Θα πρέπει να σημειωθεί ότι οι λόγοι κινδύνου  $HR_i(t)$ , για  $t=1, \dots, T$  είναι ανεξάρτητοι. Για το λόγο αυτό το συνολικό Log Hazard Ratio υπολογίζεται από ένα σταθμισμένο άθροισμα των μεμονωμένων εκτιμήσεων του Log Hazard Ratio κατά την διάρκεια του κάθε χρονικού διαστήματος  $(t-1, t)$  όπου όπου οι σταθμίσεις είναι αντιστρόφως ανάλογες της διακύμανσης κάθε εκτίμησης όπως φαίνεται και στη σχέση 22. Με τον ίδιο τρόπο προκύπτει αντίστοιχα και η διακύμανση από τη σχέση 23.

$$\ln(HR_i) = \frac{\sum_{t=1}^T \frac{\ln(HR_i(t))}{var[\ln(HR_i(t))]} }{\sum_{t=1}^T \frac{1}{var[\ln(HR_i(t))]} } \quad (22)$$

$$var[\ln(HR_i(t))] = \left[ \sum_{t=1}^T \frac{1}{var[\ln(HR_i(t))]} \right]^{-1} \quad (23)$$

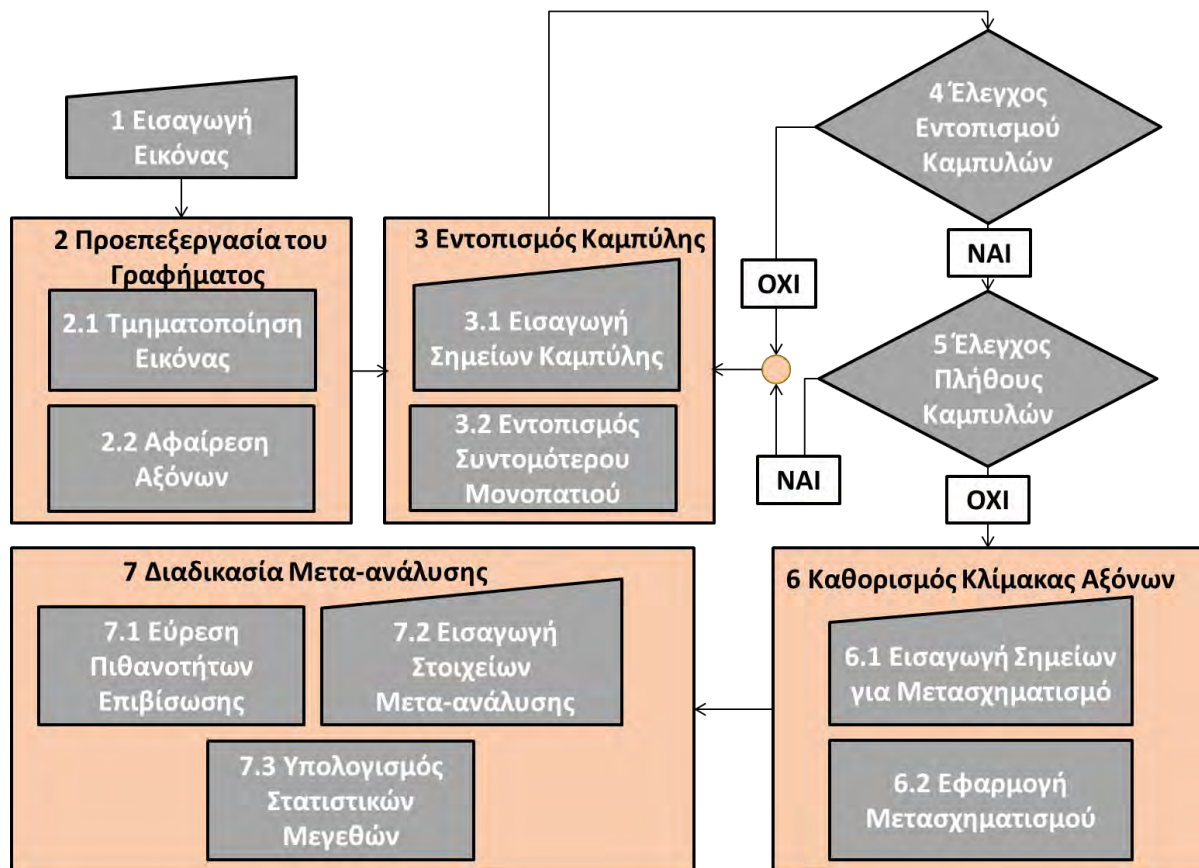
Γενικά παρουσιάζονται αρκετά προβλήματα στον υπολογισμό του Log Hazard Ratio από μελέτες που έχουμε τις δημοσιευμένες καμπύλες επιβίωσης. Ένα σημαντικό πρακτικό θέμα είναι ότι οι δημοσιευμένες καμπύλες είναι συχνά πολύ μικρές για να είναι δυνατή η ακριβής ανάγνωση ακόμα και σε περίπτωση που εφαρμοστεί διεύρυνση αυτών [16]. Η επιλογή του μήκους των χρονικών διαστημάτων είναι ένας σημαντικός παράγοντας σε αυτή την προσέγγιση. Για καλύτερα αποτελέσματα τα χρονικά διαστήματα θα πρέπει να επιλέγονται τέτοια ώστε ο ρυθμός των γεγονότων, κατά το χρονικό διάστημα που θα επιλεγεί να είναι σχετικά μικρός [24]. Κατά κανόνα, προτείνεται ότι το ποσοστό των γεγονότων εντός κάθε χρονικού διαστήματος δεν είναι περισσότερο από το 20 τοις εκατό εκείνων στην αρχή του χρονικού διαστήματος, αν και φυσικά αυτό μπορεί να μην είναι πάντα δυνατό. Ιδιαίτερο ενδιαφέρον παρουσιάζει το τελικό χρονικό διάστημα και αυτό θα πρέπει να επιλέγεται κατά τέτοιο τρόπο ώστε το τέλος του διαστήματος να είναι ίσο με τον εκτιμώμενο μέγιστο χρόνο παρακολούθησης ( $F_{max}$  στο παράρτημα). Σημειώστε ότι σε κάθε μελέτη τα χρονικά διαστήματα, δεν χρειάζεται να είναι ίσα και στην πραγματικότητα δεν θα έχουν γενικά ίσο μήκος, επειδή τα ποσοστά των γεγονότων είναι πιθανόν να αλλάξουν με την πάροδο του χρόνου. Ωστόσο, υπάρχει καλός λόγος τα χρονικά διαστήματα να είναι τα ίδια σε όλες τις δοκιμές της μετα-ανάλυσης για την κατασκευή της συνολικής καμπύλης επιβίωσης για τους ασθενείς υπό έρευνα και αυτούς υπό έλεγχο. Παρακάτω παρουσιάζεται και παράδειγμα με την εφαρμογή της συγκεκριμένης μεθοδολογίας.

## 2.6 Προτεινόμενη μεθοδολογία και εξαγωγή δεδομένων από καμπύλες Kaplan Meier για την πραγματοποίηση μετα-ανάλυσης

Η προτεινόμενη μεθοδολογία έχει σαν είσοδο μία καμπύλη επιβίωσης K-M σε μορφή bitmap, καθώς και τα διαθέσιμα στοιχεία που αφορούν πλήθος συμμετεχόντων και μέγιστο και ελάχιστο διάστημα παρακολούθησης. Ως έξοδο, παράγει το log hazard ratio και την διακύμανση για την τρέχουσα καμπύλη K-M. Επιπλέον συνδυάζει περισσότερες της μίας καμπύλες K-M για την εξαγωγή των στοιχείων της μετα-ανάλυσης. Τα ενδιάμεσα βήματα της προτεινόμενης μεθοδολογίας παρουσιάζονται.

1. Η επιλογή της εικόνας (γραφήματος) που περιέχει τις καμπύλες που μας ενδιαφέρουν για την μετα-ανάλυση.
2. Η προεπεξεργασία της εικόνας του γραφήματος (τμηματοποίηση και αφαίρεση πλέγματος) για την ευκολότερη επεξεργασία της.
3. Εξαγωγή της καμπύλης με χρήση αλγορίθμου του Dijkstra, αφού ο χρήστης εισάγει μικρό αριθμό οδηγών σημείων επί της καμπύλης.

4. Ο έλεγχος ορθού εντοπισμού της καμπύλης που γίνεται με μια ερώτηση από το σύστημα στο χρήστη. Σε περίπτωση αρνητικής απάντησης επιστρέφουμε στο βήμα 3 αλλιώς συνεχίζεται η διαδικασία με το επόμενο βήμα.
5. Ο έλεγχος ύπαρξης επιπλέον καμπυλών για εντοπισμό στην ίδια εικόνα σε περίπτωση θετικής απάντησης επιστρέφουμε στο βήμα 3 αλλιώς συνεχίζουμε με το βήμα 6.
6. Ο καθορισμός της κλίμακας των αξόνων που περιλαμβάνει την εισαγωγή σημείων με το ποντίκι για την δημιουργία του πίνακα μετασχηματισμού από το σύστημα συντεταγμένων της εικόνας (pixels) σε πραγματικές μονάδες (χρόνος και ποσοστό επιβίωσης).
7. Η ολοκλήρωση της μετα-ανάλυσης με τον υπολογισμό των στατιστικών μεγεθών που μας ενδιαφέρουν.



Εικόνα 13 Διάγραμμα λειτουργίας διαδικασιών εφαρμογής μετα-ανάλυσης.

## 2.6.1 Εισαγωγή εικόνας

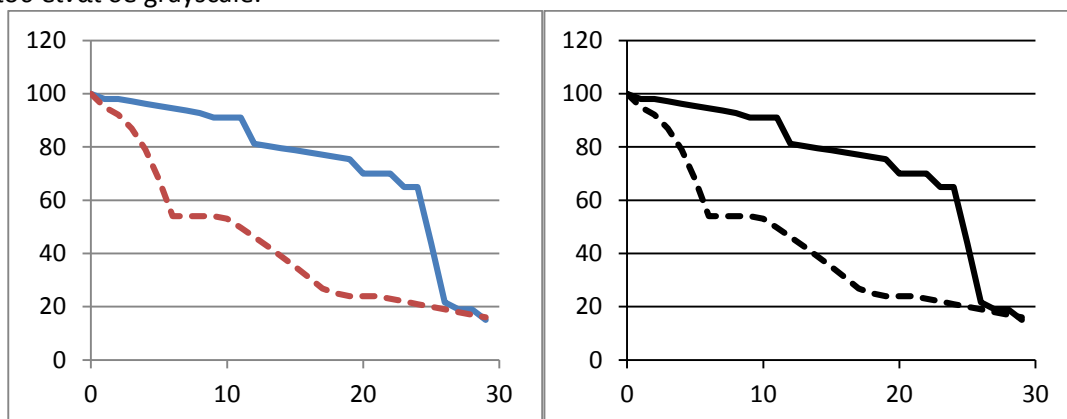
Η μέθοδος μετα-ανάλυσης που χρησιμοποιείται στην εφαρμογή είναι με άντληση στοιχείων από δημοσιευμένες εικόνες με καμπύλες Kaplan-Meier. Οι εικόνες αυτές είναι τα πιο βασικά δεδομένα που χρειάζεται η εφαρμογή για να τεθεί σε λειτουργία. Μπορεί να είναι οποιουδήποτε μεγέθους, χρώματος και μορφοποίησης. Αφού φορτωθεί η εκάστοτε εικόνα (μια κάθε φορά) στην εφαρμογή ξεκινάει η επεξεργασία της για την πραγματοποίηση της μετα-ανάλυσης.

## 2.6.2 Προεπεξεργασία του γραφήματος

το σημείο αυτό γίνεται η αφαίρεση των περισσότερων στοιχείων που δεν μας προσφέρουν κάποια πληροφορία που μπορεί να χρησιμοποιηθεί στην διαδικασία της μετα-ανάλυσης. Αυτό επιτυγχάνεται με μια σειρά από διαδικασίες και συναρτήσεις που περιγράφονται παρακάτω.

### 2.6.2.1 Τμηματοποίηση εικόνας

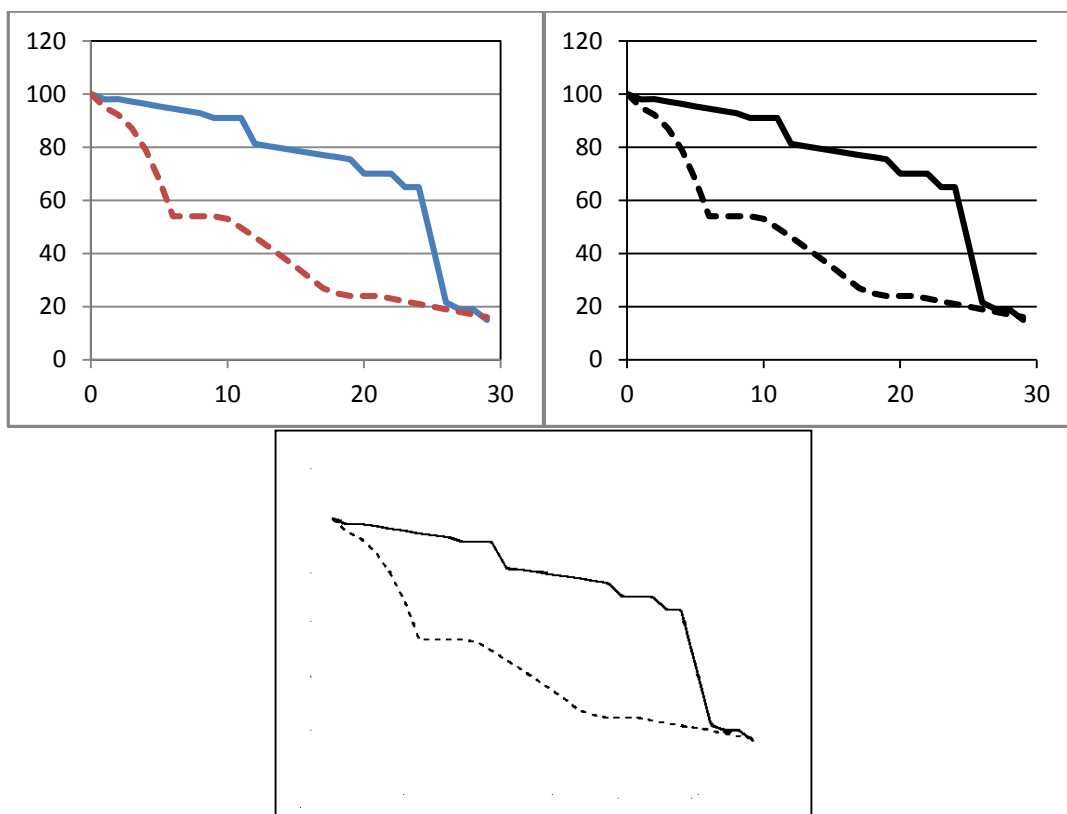
Η εικόνα του γραφήματος μετατρέπεται σε δυαδική όπως φαίνεται στην εικόνα 14. Το ίδιο ισχύει και για τις εικόνες που είναι σε grayscale.



Εικόνα 14 Η έγχρωμη φωτογραφία μετά την επεξεργασία από την `binaryImage` εμφανίζει ασπρόμαυρα όλα τα στοιχεία της ακόμα και τους άξονες.

### 2.6.2.2 Αφαίρεση αξόνων

Μια ακόμη σημαντική ενέργεια είναι η αφαίρεση κάποιων περιττών στοιχείων από την εικόνα του γραφήματος (πλέγμα, άξονες και διάφορες λεπτομέρειες) ώστε να γίνει πιο εύκολα ο εντοπισμός της στα επόμενα στάδια επεξεργασίας της. Με αυτόν τον τρόπο διευκολύνεται ο εντοπισμός μονοπατιού καθώς μειώνεται ο κίνδυνος να επιλεγεί λάθος διαδρομή. Η αφαίρεση επιτυγχάνεται με εφαρμογή του φίλτρου ενδιάμεσης (μεσαίας) τιμής. Ένα πολύ χαρακτηριστικό παράδειγμα παρουσιάζεται στην εικόνα 15 που έχουμε την γραφική παράσταση πριν και μετά την αφαίρεση των αξόνων.



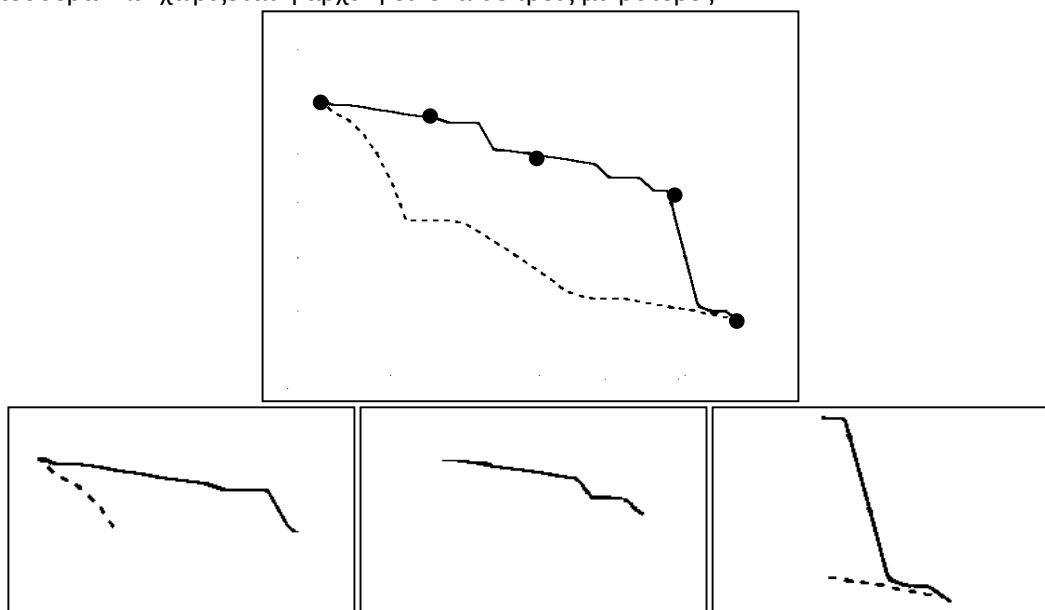
Εικόνα 15 Μια από τις συνθετικές καμπύλες πριν και μετά την επεξεργασία από τις πρώτες συναρτήσεις της εφαρμογής μας και το αποτέλεσμα.

### 2.6.3 Εντοπισμός καμπύλης

Αφού προκύψει μια εικόνα που απεικονίζει κυρίως στοιχεία της καμπύλης αρχίζει το επόμενο στάδιο επεξεργασίας που στοχεύει στον εντοπισμό αυτής με την μικρότερη δυνατή συμμετοχή του χρήστη. Για να επιτευχθεί θα εφαρμοστούν κατάλληλες μέθοδοι και αλγόριθμοι που περιγράφονται παρακάτω.

#### 2.6.3.1 Εισαγωγή σημείων καμπύλης

Αναλόγως με το πλήθος των καμπυλών και το πόσο κοντά είναι η μια στην άλλη ο χρήστης εισάγει σημεία οδηγούς στην καμπύλη ενδιαφέροντος για την καθοδήγηση του αλγορίθμου εντοπισμού συντομότερου μονοπατιού. Το πλήθος των σημείων εξαρτάται από το μήκος της καμπύλης και την (πιθανή) τομή της με άλλη καμπύλη του ίδιου γραφήματος. Το πλήθος σε κάθε περίπτωση είναι ένας μικρός μονοψήφιος αριθμός. Έτσι η συμμετοχή του χρήστη είναι μικρή και η προτεινόμενη μέθοδος είναι ιδιαίτερα πρακτική και εύρωστη. Η εικόνα χωρίζεται σε μικρότερες εικόνες που αποτελούν τμήματα της αρχικής ανάλογα με σημεία που εισήχθησαν από τον χρήστη. Στα συγκεκριμένα τμήματα εφαρμόζεται ξεχωριστά ο αυτόματος εντοπισμός του συντομότερου μονοπατιού και στο τέλος εμφανίζεται το συνολικό αποτέλεσμα. Στην εικόνα 14 φαίνεται πως κάνοντας τέσσερα κλικ χωρίζεται η αρχική εικόνα σε τρεις μικρότερες.



Εικόνα 16 Αρχική εικόνα προς επεξεργασία και τα τμήματα της αυτής όπως προέκυψαν από τα κλικ του χρήστη για την καμπύλη με συνεχόμενη γραμμή.

#### 2.6.3.2 Εντοπισμός συντομότερου μονοπατιού

Το βήμα αυτό περιλαμβάνει την ανίχνευση της καμπύλης στην εικόνα που επεξεργαζόμαστε και βασίζεται στην χρήση ενός αλγορίθμου εύρεσης συντομότερου μονοπατιού που θα παρουσιαστεί με λεπτομέρειες στη συνέχεια.

Το πρόβλημα εύρεσης του βέλτιστου μονοπατιού σε ένα κατευθυνόμενο γράφο  $G$  με βάρη  $V$  και ακμές  $E$ :  $G=(V,E)$  έχει μελετηθεί εκτενώς (πχ.)  $G=(V,E)$  [25]. Η εικόνα του γραφήματος  $K-M$  είναι ένας μεγάλος πίνακας που αποτελείται από κελιά (pixels) με γνωστές τιμές. Στην συγκεκριμένη περίπτωση τον ρόλο των κόμβων του γράφου παίζουν τα κελιά και τον ρόλο των ακμών με κόστος μετακίνησης από τον ένα κελί στο άλλο οι τιμές τους. Κατασκευάζουμε μια συνάρτηση  $w: E \rightarrow R$  που καταγράφουμε όλες τις πραγματικές τιμές των ακμών. Το μήκος του μονοπατιού  $p = \{u_0, u_1, \dots, u_k\}$  ορίζεται σαν το άθροισμα των βαρών των συνδέσεων  $w$  των pixels που συμμετέχουν στη δημιουργία του.

$$w(p) = \sum_{i=1}^k w(u_{i-1}, u_i)$$

με την προϋπόθεση να υπάρχει μονοπάτι από το  $u$  στο  $v$  αλλιώς είναι άπειρη απόσταση.  $uvrw(p) = \delta(u, v)$

Ο αλγόριθμος του Dijkstra λύνει το πρόβλημα συντομότερου μονοπατιού μιας πηγής σε ένα κατευθυνόμενο γράφο με βάρη  $G = (V, E)$  για την περίπτωση κατά την οποία όλα τα βάρη των ακμών είναι μη αρνητικά. Ο αλγόριθμος τηρεί τις εξής κύριες δομές δεδομένων:

- έναν πίνακα  $S$  ο οποίος κρατά τους κόμβους του μονοπατιού που έχουν βρεθεί την τρέχουσα χρονική στιγμή
- έναν πίνακα  $d$ , μεγέθους ίσου με το πλήθος όλων των κόμβων του γράφου (ίσο με τον πληθάρημο του  $V$ ) ο οποίος τηρεί το κόστος του μονοπατιού από το αρχικό σημείο έως τον κάθε κόμβο
- τον πίνακα των γονέων,  $\pi$ , μεγέθους ίσου με το πλήθος όλων των κόμβων του γράφου, ο οποίος κρατά τον προηγούμενο κόμβο του βέλτιστου μονοπατιού από τον αρχικό έως τον τρέχοντα κόμβο.

Πολύ σημαντική στους αλγορίθμους εύρεσης ελάχιστου μονοπατιού είναι η τεχνική της «χαλάρωσης». Έστω ότι έχει βρεθεί ένα βέλτιστο μονοπάτι από τον αρχικό κόμβο  $u_0$  στον  $u$ . Έστω  $v$  κόμβος συνδεδεμένος με τον  $u$ , ο οποίος δεν περιλαμβάνεται στο τρέχον μονοπάτι.  $u, v$ . Η διαδικασία της χαλάρωσης της ακμής  $(u, v)$ ,  $(u, v)$  ελέγχει αν μπορεί να βελτιωθεί το μονοπάτι που έχει βρεθεί μέχρι τώρα από το  $u_0$  στο  $u$ , συμπεριλαμβάνοντας το  $v$ . Αν το νέο μονοπάτι είναι βέλτιστο, συμπεριλαμβάνεται ο κόμβος  $v$  *υναλλάζοντας* τα  $d(u)$  και  $\pi(u)$ . Η εκτέλεση της χαλάρωσης μπορεί να μειώσει το κόστος του το συντομότερου μονοπατιού  $d(u)$  και μετά να ανανεώσει το πεδίο  $\pi(u)$ . Ακολουθεί ψευδοκώδικας με εφαρμογή της συνάρτησης χαλάρωσης σε μια ακμή  $(u, v)$ .

#### ΣΥΝΑΡΤΗΣΗ-ΧΑΛΑΡΩΣΗΣ( $u, v, w$ )

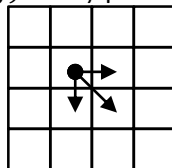
- 1 **Αν**  $d(u) > d(u) + w(u, v)$
- 2     **Τότε**  $d(u) \leftarrow d(u) + w(u, v)$
- 3      $\pi(u) \leftarrow \pi(u) \text{ sud}(u)$

Ο αλγόριθμος του Dijkstra αρχικοποιεί το κόστος κάθε κόμβου ως προς τον αρχικό ίσο με μία μέγιστη τιμή (εκτός από την τιμή που αφορά τον αρχικό κόμβο) και τον πίνακα των γονέων σε τιμή Null και ως εξής :

#### ΣΥΝΑΡΤΗΣΗ-ΑΡΧΙΚΟΠΟΙΗΣΗΣ-ΜΟΝΗΣ-ΔΙΑΔΡΟΜΗΣ ( $G, s$ )

- 1 **Για** κάθε κόμβο  $u \in V$
- 2     **Υλοποίησε**  $d(u) \leftarrow \infty$
- 3      $\pi(u) \leftarrow NULL$
- 4  $d(s) \leftarrow 0$

Έτσι μετά την αρχικοποίηση [25] του  $\pi(u) = NULL$  για κάθε  $u \in V$ , και  $d(u) = 0$  για  $u = s$  και  $d(u) = \infty$  για κάθε  $u \in V - \{s\}$ .  $G = (V, E)$ . Στην περίπτωση της εφαρμογής της συγκεκριμένης εργασίας οι καμπύλες οι οποίες πρέπει να ακολουθηθούν είναι φθίνουσες, εφόσον πρόκειται για μέτρηση επιβίωσης συναρτήσε του χρόνου. Σε ένα γράφο κατά την εκτέλεση του αλγορίθμου γίνεται επίσκεψη των γειτονικών κόμβων ενώ στην περίπτωση της εφαρμογής αν εξετάζουμε την απόσταση από το κελί  $(i, j)$  πραγματοποιείται επίσκεψη μόνο στα κελιά  $(i, j + 1)$  και  $(i + 1, j)$  όπως φαίνεται στο επόμενο σχήμα:



Εικόνα 17 Οι συνδεδεμένοι γειτονικοί κόμβοι ενός τρέχοντος κόμβου (σημασμένου με τελεία) στην περίπτωση της παρούσας εφαρμογής.

Δεδομένου ενός κατευθυνόμενου γράφου με βάρη  $G = (V, E)$ , υποθέτουμε ότι το  $w(u, v) \geq 0$  για κάθε ακμή  $(u, v) \in E$ . Όπως αναφέρθηκε, ο Αλγόριθμος του Dijkstra διατηρεί ένα σύνολο κόμβων  $S$  του οποίου η τελική συντομότερη διαδρομή από την πηγή  $s$  έχει ήδη προσδιοριστεί. Αυτό ισχύει για όλους  $u \in S$  και έχουμε  $d(u) = \delta(s, u)$ . Ο αλγόριθμος επιλέγει κατ' επανάληψη έναν κόμβο από τον  $u \in V - S$  με υπολογισμένο το μήκος του ελάχιστου μονοπατιού, εισάγει το  $u$  στο  $S$  και χαλαρώνει όλες τις ακμές του  $u$ . Ακολουθεί ο ψευδοκώδικας λειτουργίας του αλγορίθμου στον οποίο χρησιμοποιούμε μια σειρά προτεραιότητας  $Q$  που περιέχει όλους τους κόμβους  $V - S$  τοποθετημένους ανάλογα με την  $d$  τιμή τους.

## ΑΛΓΟΡΙΘΜΟΣ-DIJKSTRA

```
1 ΣΥΝΑΡΤΗΣΗ-ΑΡΧΙΚΟΠΟΙΗΣΗΣ-ΜΟΝΗΣ-ΔΙΑΔΡΟΜΗΣ ( $G, s$ )
2  $S \leftarrow \emptyset$ 
3  $Q \leftarrow V[G]$ 
4 Όσο  $Q \neq \emptyset$ 
5   υπολόγισε  $u \leftarrow$  κόμβος του Q-S με ελάχιστο κόστος ( $Q$ )
6    $S \leftarrow S \cup \{u\}$ 
7   Για κάθε κόμβο  $v$  γειτονικό του  $uv \in V$ 
8     Κάλεσε ΣΥΝΑΡΤΗΣΗ-ΧΑΛΑΡΩΣΗΣ( $u, v, w$ )
```

Η πολυπλοκότητα του αλγορίθμου υπολογίζεται ως εξής. Με δεδομένο ότι το Q αρχικοποιείται με όλες τις κορυφές του γράφου, ο βρόχος στις γραμμές 4 – 8 επαναλαμβάνεται τόσες φορές όσες είναι οι κορυφές του γράφου (ισοδύναμα τα pixels της εικόνας).  $Q = V - S$   $O(V)$ . Η αναζήτηση του κόμβου  $u$  στην γραμμή 5 θα χρειαστεί  $O(V^2)$ . Κάθε κόμβος  $u \in V$  εισάγεται στο σύνολο  $S$  μια φορά ακριβώς, έτσι για κάθε ακμή της εικόνας-πίνακα εξετάζεται μια φορά κατά την διάρκεια λειτουργίας του αλγορίθμου. Αφού ισχύει ότι ο συνολικός αριθμός των ακμών είναι όλα τα στοιχεία  $E$  της εικόνας-πίνακα πραγματοποιούνται  $E$  επαναλήψεις με την κάθε μία να θέλει χρόνο ολοκλήρωσης  $O(1)$ . Ο συνολικός χρόνος για την ολοκλήρωση του αλγορίθμου είναι  $O(V^2 + E) = O(V^2)$ .

Η συνάρτηση χαλάρωσης ενημερώνει το πίνακα των γονέων του κάθε κόμβου του γράφου. Μετά την ολοκλήρωση της εκτέλεσης, το μονοπάτι βρίσκεται ξεκινώντας από τον τελικό κόμβο και κινούμενοι προς τον αρχικό, χρησιμοποιώντας τις τιμές του γονέα του κάθε κόμβου.

Μετά την ολοκλήρωση εκτέλεσης του αλγορίθμου του Dijkstra έχουμε σαν αποτέλεσμα μια κόκκινη γραμμή που απεικονίζει την καμπύλη πάνω στην αρχική εικόνα με σκοπό να πραγματοποιηθεί η διαδικασία ελέγχου εντοπισμού της καμπύλης.

### 2.6.4 Έλεγχος εντοπισμού καμπύλης

Το τελικό αποτέλεσμα της όλης διαδικασίας εμφανίζεται στο χρήστη σε μια εικόνα και μέσω ενός dialog box του Matlab ο χρήστης απαντάει για το αν έχει εντοπιστεί σωστά η καμπύλη. Σε περίπτωση που δεν έχει εκτελεσθεί σωστά ο εντοπισμός της καμπύλης η διαδικασία επιστρέφει στην αρχή της συγκεκριμένης διαδικασίας με τον χρήστη να ξανακάνει κλικ όπως παρουσιάστηκε παραπάνω μέχρι να προκύψει το επιθυμητό αποτέλεσμα.

### 2.6.5 Έλεγχος πλήθους καμπυλών

Μετά την ερώτηση αν έχει γίνει σωστά ο εντοπισμός της καμπύλης από την εφαρμογή εμφανίζεται ένα ακόμα dialog box στο οποίο ο χρήστης απαντάει για το αν υπάρχουν άλλες καμπύλες προς ανίχνευση. Σε περίπτωση που υπάρχουν και άλλες καμπύλες η εφαρμογή επιστρέφει στο βήμα εντοπισμού της καμπύλης μέχρι τελικά να βρεθούν όλες. Όταν τελειώσει ο εντοπισμός όλων των καμπυλών που απεικονίζονται στην εικόνα που έχει τεθεί προς επεξεργασία ο αλγόριθμος προχωράει στο επόμενο βήμα.

### 2.6.6 Καθορισμός κλίμακας αξόνων

Αφού έχουν ολοκληρωθεί σωστά οι προηγούμενες διεργασίες και ενέργειες της εφαρμογής η όλη διαδικασία προχωράει στον καθορισμό της κλίμακας των αξόνων. Σκοπός του βήματος αυτού είναι η κατασκευή του μετασχηματισμού από το σύστημα συντεταγμένων του γραφήματος (pixels) στις πραγματικές μονάδες των σημείων της καμπύλης (χρόνος και ποσοστό επιβίωσης). Ο χρήστης πρέπει να δώσει τρία τουλάχιστον σημεία για να είναι δυνατή η κατασκευή του πίνακα μετασχηματισμού. Ο μετασχηματισμός που θα εφαρμόσουμε ονομάζεται συσχετισμένος μετασχηματισμός (Affine Transformation). Ο μετασχηματισμός αυτός μπορεί να χειριστεί οποιοδήποτε συνδυασμό γεωμετρικών μετασχηματισμών που περιλαμβάνουν αλλαγή κλίμακας, μετατόπιση και περιστροφή.

Το συγκεκριμένο στάδιο περιλαμβάνει την εισαγωγή σημείων των αξόνων με κλικ από τον χρήστη μετά την εμφάνιση του μηνύματος 'insert transform points'. Ο χρήστης πρέπει να δώσει τουλάχιστον 3 σημεία ή παραπάνω για μεγαλύτερη ακρίβεια, οι τιμές των οποίων αποθηκεύονται σε μεταβλητές θα χρησιμοποιηθούν από συνάρτηση της εφαρμογής και περιέχουν τις τιμές των pixel της εικόνας. Για κάθε ένα από αυτά τα σημεία που ουσιαστικά είναι pixel της εικόνας, ο χρήστης πρέπει να εισάγει τις πραγματικές τιμές που αντιστοιχεί σ' αυτά δηλαδή την πιθανότητα επιβίωσης και τον χρόνο. Οι πραγματικές τιμές των σημείων δίνονται από το πληκτρολόγιο σε ένα dialog box που εμφανίζεται στην οθόνη με το κατάλληλο μήνυμα. Με τα σημεία αυτά θα βρεθεί ο μετασχηματισμός που εφαρμόστηκε για να προκύψουν αυτές οι αντιστοιχίες.

Έστω  $N$  σημεία επί των 2 αξόνων του γραφήματος με συντεταγμένες στην εικόνα  $\{(i_k, j_k)\}$  και πραγματικές τιμές χρόνου και επιβίωσης  $\{(x_k, y_k)\}$ . Αναζητούμε τον συσχετισμένο μετασχηματισμό (Affine Transformation) με πίνακα σε ομογενείς συντεταγμένες ο οποίος ελαχιστοποιεί την ποσότητα  $\sum_{k=1}^N (z_k - y_k)^2 = \min$ , όπου  $(v_k, z_k, 1)^T = M((j_k, z_k, 1)^T)$ .

$$M = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ 0 & 0 & 1 \end{bmatrix}$$

Για να καθορίσουμε τον μετασχηματισμό Affine χρειαζόμαστε τον πίνακα του μετασχηματισμού ο οποίος έχει 6 αγνώστους τα  $a_{00}, a_{01}, a_{02}, a_{10}, a_{11}, a_{12}$ . Οι 3 άγνωστοι  $a_{00}, a_{01}, a_{02}$  υπολογίζονται από τις  $X$  συντεταγμένες των ζευγών ομολόγων σημείων και οι 3 άγνωστοι  $a_{10}, a_{11}, a_{12}$  υπολογίζονται από τις  $Y$  συντεταγμένες. Αν το πλήθος των ζευγών σημείων είναι  $N$  πρέπει να επιλυθούν 2 γραμμικά συστήματα  $N \times N$ . Αν όπως συνήθως έχουμε  $N > 3$  τότε έχουμε 2 γραμμικά συστήματα για το καθένα από τα οποία έχουμε 3 αγνώστους και  $N > 3$  γραμμικές εξισώσεις. Αν ορίσουμε

$$A = \begin{bmatrix} j_1 & i_1 & 1 \\ j_2 & i_2 & 1 \\ \vdots & \vdots & \vdots \\ j_N & i_N & 1 \end{bmatrix}$$

$$B_x = [x_1, x_2, \dots, x_N]^T, B_y = [y_1, y_2, \dots, y_N]^T$$

τότε τα 2 γραμμικά συστήματα που πρέπει να επιλυθούν είναι

$$\begin{aligned} A[a_{00}, a_{01}, a_{02}]^T &= B_x \\ A[a_{10}, a_{11}, a_{12}]^T &= B_y \end{aligned}$$

Στην περίπτωση που τα συστήματα αυτά είναι υπερκαθορισμένα ( $N > 3$ ), τότε η επίλυση γίνεται με την μέθοδο των ελαχίστων τετραγώνων, ως εξής:

$$\begin{aligned} [a_{00}, a_{01}, a_{02}]^T &= (A^T A)^{-1} B_x \\ [a_{10}, a_{11}, a_{12}]^T &= (A^T A)^{-1} B_y \end{aligned}$$

Η υλοποίηση των προηγούμενων σχέσεων σε περιβάλλον Matlab χρησιμοποιούμε τον τελεστή «\»..Αφού έχει υπολογιστεί ο πίνακας  $M$ , το σύνολο των διατεταγμένων ζευγών pixel που επιστρέφει ο αλγόριθμος του Dijkstra μετατρέπονται σε ζεύγη χρόνου επιβίωσης, πολλαπλασιαζόμενα με τον πίνακα  $M$ , όπως αναφέρθηκε παραπάνω.



### 2.6.6.1 Εφαρμογή του Μετασχηματισμού

Αφού υπολογιστεί ο πίνακας μετασχηματισμού εφαρμόζεται σε όλα σημεία της καμπύλης για να υπολογιστούν οι πραγματικές τιμές της. Οι τιμές αυτές είναι οι πιθανότητες επιβίωσης και ο χρόνος που αντιστοιχεί σε αυτές. Ο κώδικας για τον υπολογισμό του πίνακα μετασχηματισμού και την εφαρμογή του παρουσιάζεται στο παράρτημα. Μετά την ολοκλήρωση των διαδικασιών καθορισμού της κλίμακας των αξόνων σειρά έχει το στάδιο της μετα-ανάλυσης.

### 2.6.7 Διαδικασία Μετα-ανάλυσης

Κατά την εκτέλεση του συγκεκριμένου βήματος πραγματοποιείται ο υπολογισμός των μεγεθών και μεταβλητών που έχουν σχέση με την μετα-ανάλυση. Πριν ξεκινήσει η λειτουργία της εφαρμογής ο χρήστης έχει ορίσει τα χρονικά διαστήματα σε ένα αρχείο txt το οποίο θα χρησιμοποιηθεί για να ανακτηθούν τα δεδομένα αυτά που θα χρειαστούν για την πραγματοποίηση του συγκεκριμένου βήματος. Το πρώτο χρονικό διάστημα είναι το  $(0, t_i)$  και αυτό που υπολογίζεται είναι η πιθανότητα επιβίωσης στην αρχή του κάθε διαστήματος. Έτσι για  $t_{si} = 0$  που είναι η αρχή του πρώτου χρονικού διαστήματος εξορισμού έχουμε 100% πιθανότητα επιβίωσης. Για κάθε ένα από τα επόμενα χ.δ. πραγματοποιείται ο υπολογισμός της πιθανότητας στην οποία αντιστοιχεί.

Πραγματοποιείται η εισαγωγή μεγεθών από τον χρήστη μέσω του πληκτρολογίου που αφορούν το πλήθος των συμμετεχόντων στη μελέτη και την μέγιστη και την ελάχιστη διάρκεια παρακολούθησης της μελέτης.

Τα στοιχεία που είναι απαραίτητα για τον υπολογισμό του Log Hazard Ratio και της διακύμανσης έχουν πλέον όλα συλλεχθεί και χρησιμοποιούνται από το σύστημα με βάση τις σχέσεις που παρουσιάστηκαν στο κεφάλαιο 2 που περιγράφεται η μεθοδολογία της μετα-ανάλυσης που χρησιμοποιήθηκε για την ανάπτυξη της εφαρμογής.

### 3 Υλικά, υπολογιστική υλοποίηση

#### 3.1 Διαθέσιμα δεδομένα

Για την εύρεση βιβλιογραφίας πραγματοποιήθηκε έρευνα σε δημοσιευμένες εργασίες και μελέτες σε περιοδικά που βρέθηκαν στο διαδίκτυο. Οι συγκεκριμένες πηγές έχουν σαν αντικείμενο τις αναλύσεις επιβίωσης και πιο συγκεκριμένα την μετα-ανάλυση ή άλλα επιμέρους θέματα που έχουν σχέση με τον κώδικα που γράφτηκε για την δημιουργία της εφαρμογής. Κατά την αναζήτηση για πληροφορίες και θέματα σε σχέση με την μετα-ανάλυση βρέθηκε παρόμοια εφαρμογή με αυτή της παρούσας εργασίας όπου πραγματοποιήθηκε ποιοτική και ποσοτική σύγκριση όπως θα παρουσιαστεί στο κεφάλαιο 4.

Από την συνολική βιβλιογραφία χρησιμοποιήθηκαν κομμάτια που περιείχαν αλγορίθμους εύρεσης μονοπατιών, εφαρμογή μετασχηματισμών, διάφορες μεθόδους επεξεργασίας εικόνas και πληροφορίες για την χρήση του εργαλείου Matlab. Επίσης όσον αφορά το κομμάτι της μετα-ανάλυσης οι εργασίες που επιλέξαμε είχαν αρκετές πληροφορίες για την αξιοποίηση δεδομένων από καμπύλες Kaplan Meier που είναι χρήσιμες και απαραίτητες για τον υπολογισμό των στατιστικών μεγεθών όπως της διακύμανση και του log hazard ratio.

Σκοπός της συγκεκριμένης εργασίας είναι η δημιουργία μιας εφαρμογής που να ενσωματώνει την μεθοδολογία που παρουσιάστηκε στο κεφάλαιο 2 για την πραγματοποίηση μιας μετα-ανάλυσης από δημοσιευμένες καμπύλες Kaplan Meier. Η εφαρμογή αυτή δημιουργήθηκε σε περιβάλλον Matlab χρησιμοποιώντας μια σειρά αλγορίθμων καθώς επίσης στοιχεία και σχέσεις που συνδέονται με την διαδικασία της μετα-ανάλυσης για να φτάσουμε στο επιθυμητό αποτέλεσμα.

#### 3.2 Περιβάλλον ανάπτυξης Matlab

Το MATLAB αποτελεί ένα περιβάλλον για την πραγματοποίηση αριθμητικών υπολογισμών, την δημιουργία διαφόρων απεικονίσεων και την χρήση σαν γλώσσα προγραμματισμού υψηλού επιπέδου. Χρησιμοποιώντας το MATLAB μπορούμε να αναλύσουμε δεδομένα, να αναπτύξουμε αλγορίθμους καθώς επίσης και να δημιουργήσουμε εφαρμογές και διάφορα μοντέλα [26]. Πρόκειται για ένα λογισμικό που προσφέρει πολλά εργαλεία, έτοιμες συναρτήσεις και μια γλώσσα προγραμματισμού που σε συνδυασμό με όλα τα υπόλοιπα δίνουν στο χρήστη την δυνατότητα να αναλύσει και να ψάξει ένα θέμα σε βάθος για να καταλήξει σε συμπεράσματα και λύσεις πολύ ταχύτερα από ότι χρησιμοποιώντας τις παραδοσιακές μεθόδους.

Το MATLAB μπορεί να χρησιμοποιηθεί για ένα ευρύ φάσμα εφαρμογών όπως επεξεργασία σήματος και επικοινωνιών, επεξεργασία εικόνas και βίντεο, συστήματα ελέγχου, δοκιμών και μετρήσεων, υπολογιστική οικονομία και υπολογιστική βιολογία. Η έκδοση που χρησιμοποιήθηκε για την ανάπτυξη της δικιάς μας εφαρμογής που έχει σχέση με στατιστικά μεγέθη της βιολογίας είναι η R2008b.

#### 3.3 Παρουσίαση της εφαρμογής

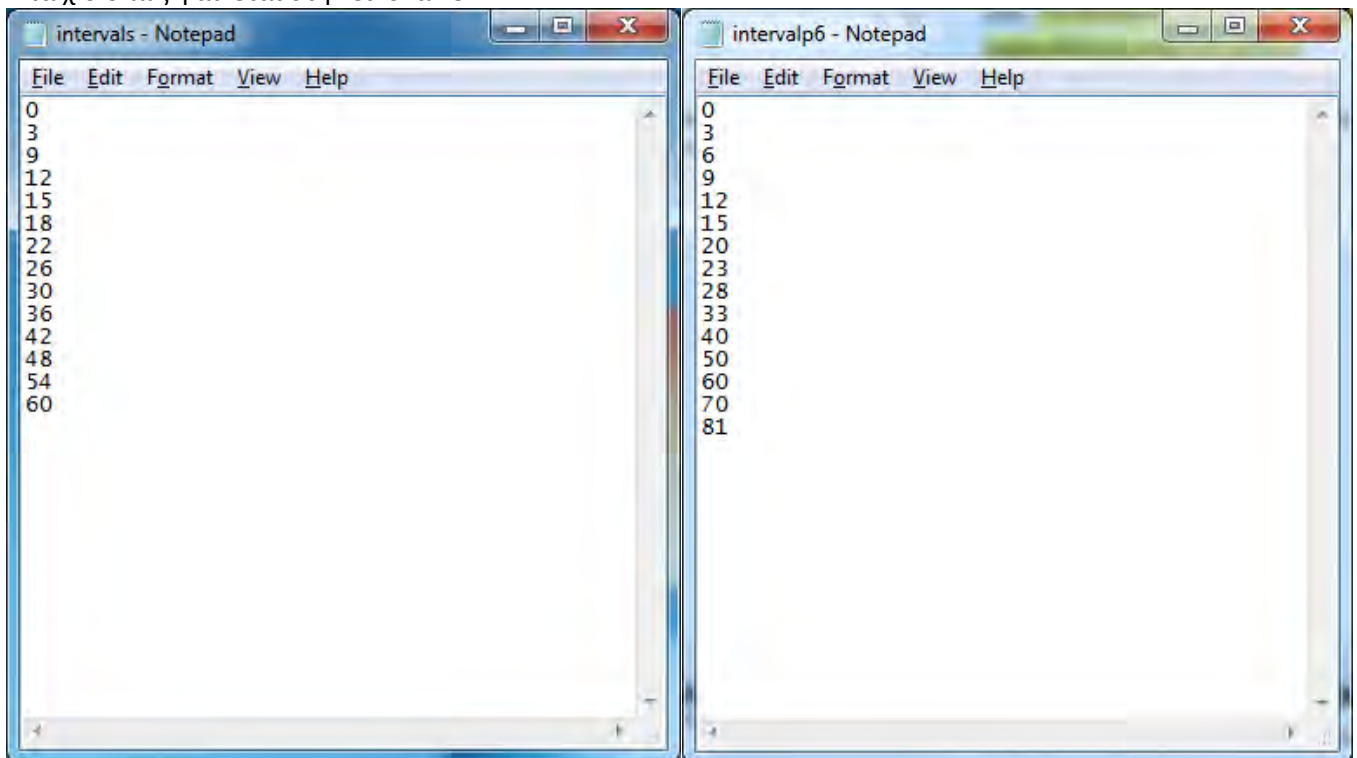
Η εφαρμογή που κατασκευάστηκε χρησιμοποιεί δεδομένα όπως εικόνες όλων των μορφοποιήσεων και αρχεία κειμένου (txt) από όπου αντλεί δεδομένα για την πραγματοποίηση μετα-ανάλυσης. Τα αρχεία αυτά πρέπει να βρίσκονται στο ίδιο directory του υπολογιστή στον οποίο θα γίνεται και η αποθήκευση των δεδομένων εξόδου μετά την εκτέλεση της εφαρμογής μας. Τα δεδομένα εξόδου αποθηκεύονται σε αρχείο excel για περαιτέρω επεξεργασία η αρχειοθέτηση από τον χρήστη.

#### 3.4 User interface

Για να τεθεί σε λειτουργία η εφαρμογή της παρούσας εργασίας ο χρήστης θα πρέπει να ανοίξει το αρχείο MetanAlysis.fig. Για την πραγματοποίηση της μετα-ανάλυσης θα πρέπει να εκτελεσθούν ορισμένα βήματα με συγκεκριμένη σειρά για να έχουμε τα επιθυμητά αποτελέσματα.

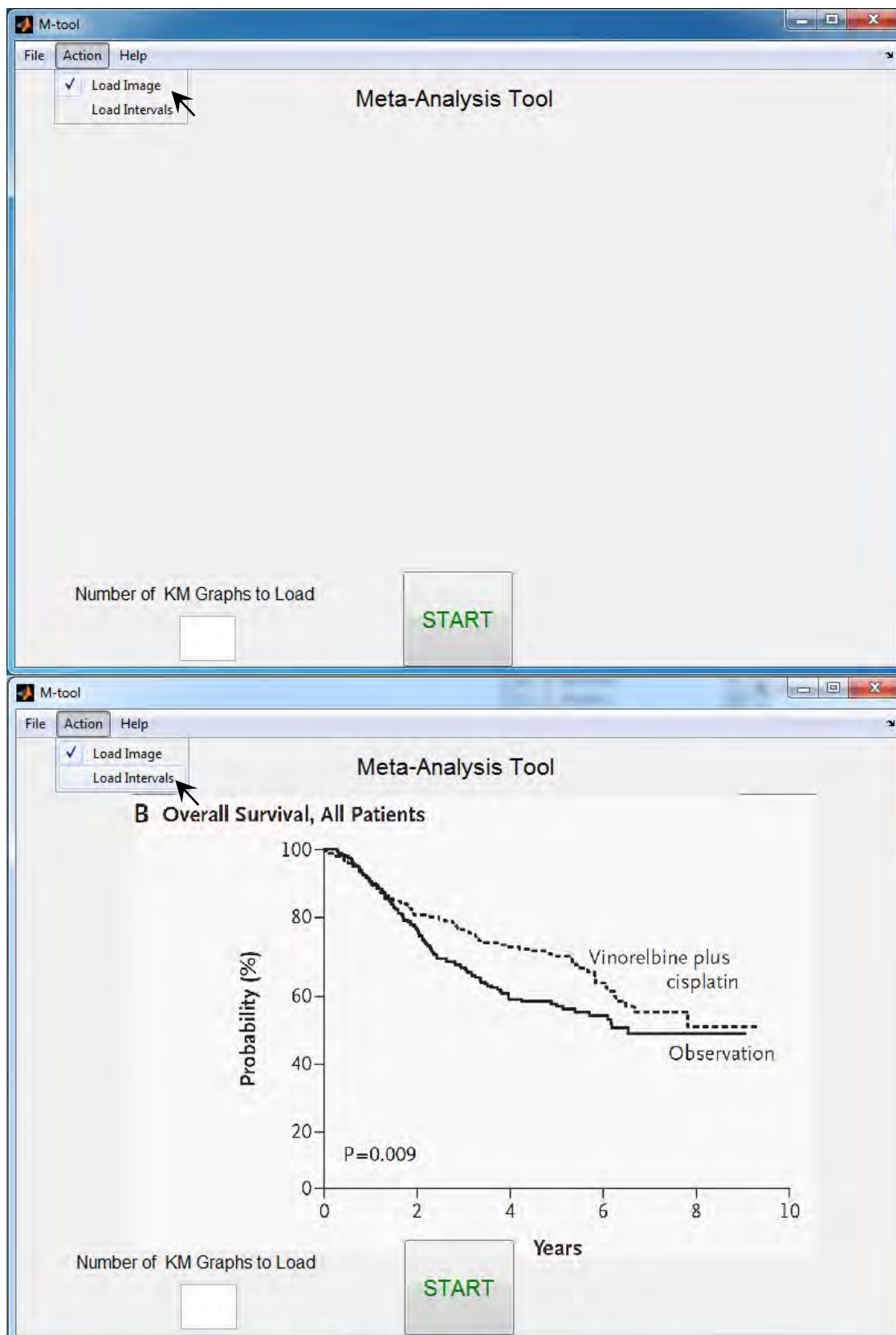
Η πρώτη ενέργεια πριν την εκτέλεση της εφαρμογής είναι ο ορισμός των χρονικών σημείων σε ένα αρχείο txt όπου γράφεται το χρονικό σημείο από την έναρξη μέχρι και την λήξη της μελέτης ορισμένα με κατάλληλο τρόπο που περιγράφεται στο υποκεφάλαιο 2.2 της μεθοδολογίας. Το πρώτο χρονικό σημείο είναι το 0, και το τελευταίο είναι ο συνολικός χρόνος για την ολοκλήρωση της μελέτης. Ο κάθε αριθμός που αντιπροσωπεύει ένα διαφορετικό χρονικό σημείο γράφεται σε διαφορετική γραμμή (μορφή πίνακα στήλη) μέσα στο ίδιο αρχείο txt. Τα χρονικά σημεία είναι ουσιαστικά μήνες και ανά δύο διαδοχικά σχηματίζουν το χρονικό διάστημα που αναζητούμε τα συγκεκριμένα στατιστικά μεγέθη της μετα-ανάλυσης. Η μορφή του αρχείου φαίνεται στην εικόνα 17 για μια μελέτη που διήρκεσε 5 χρόνια και μια άλλη που διήρκεσε 7 σχεδόν χρόνια.

Το αρχείο αυτό πρέπει να είναι στον ίδιο φάκελο που βρίσκονται η εφαρμογή και τα υπόλοιπα δεδομένα που θα χρησιμοποιηθούν. Θα γίνει η επιλογή του μέσα από την εφαρμογή για να χρησιμοποιηθούν τα στοιχεία που περιέχει. Η ενέργεια αυτή θα εκτελεσθεί με την επιλογή File και στην συνέχεια την επιλογή Load intervals που θα εμφανιστεί ένα παράθυρο για να γίνει εντοπισμός και επιλογή του αρχείου που περιέχει τα χ.δ όπως φαίνεται στην εικόνα 19.



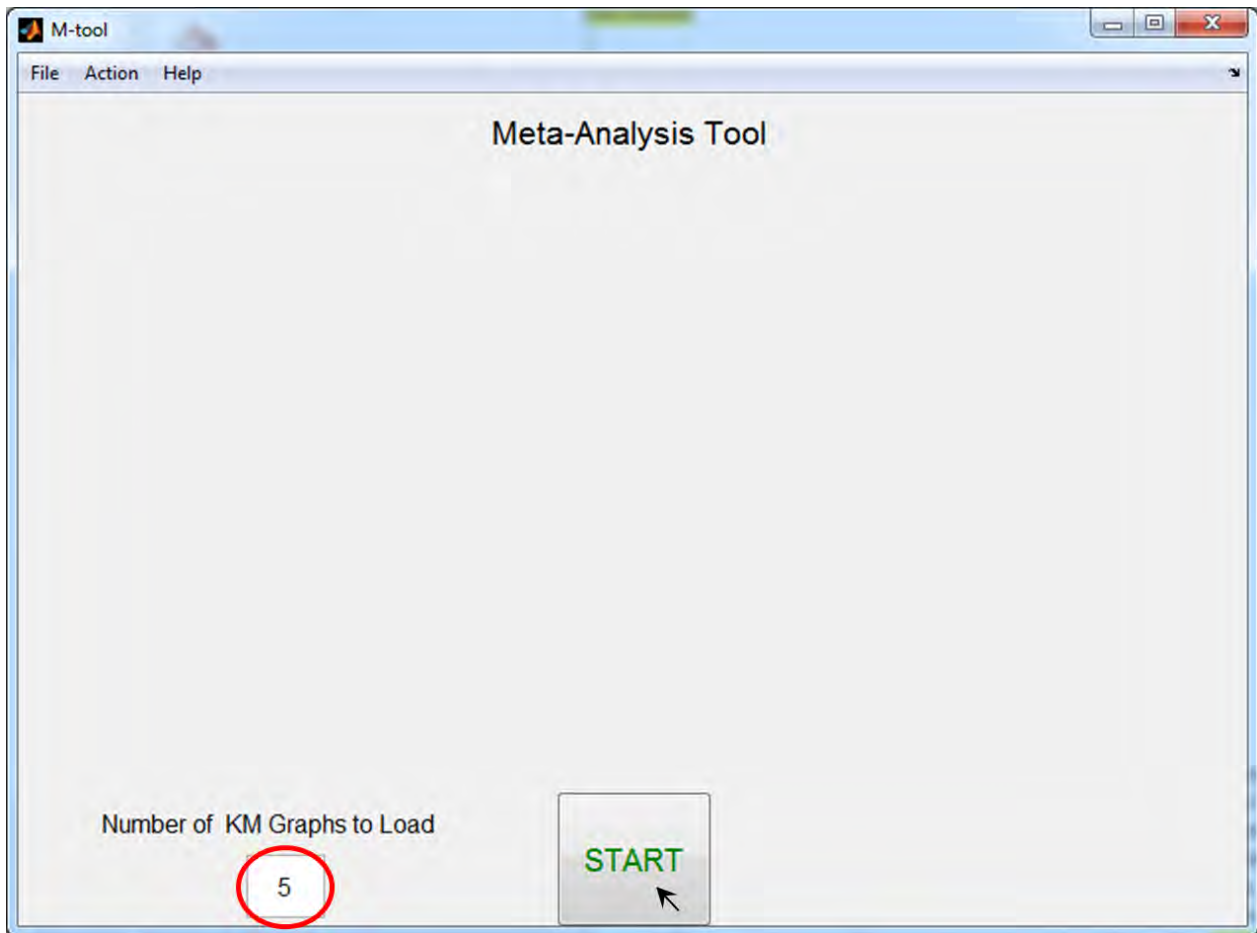
Εικόνα 18 Παραδείγματα δημιουργίας χρονικών σημείων σε αρχεία txt που αναπαριστούν μήνες και ανά δύο σχηματίζουν χρονικά διαστήματα.

Επόμενο βήμα είναι η επιλογή μιας συγκεκριμένης εικόνας που περιέχει αποτελέσματα από κάποια μελέτη και θα χρησιμοποιηθεί για την πραγματοποίηση μετα-ανάλυσης με την βοήθεια της εφαρμογής. Η δυνατότητα επιλογής βρίσκεται στο μενού πατώντας αρχικά το File και επιλέγοντας στην συνέχεια την εντολή Load Image. Ακολουθώντας τα παραπάνω βήματα θα εμφανιστεί ένα παράθυρο με τις εικόνες που περιέχει το directory που βρίσκονται και τα υπόλοιπα στοιχεία της εφαρμογής. Το παραπάνω βήμα φαίνεται στην εικόνα 20

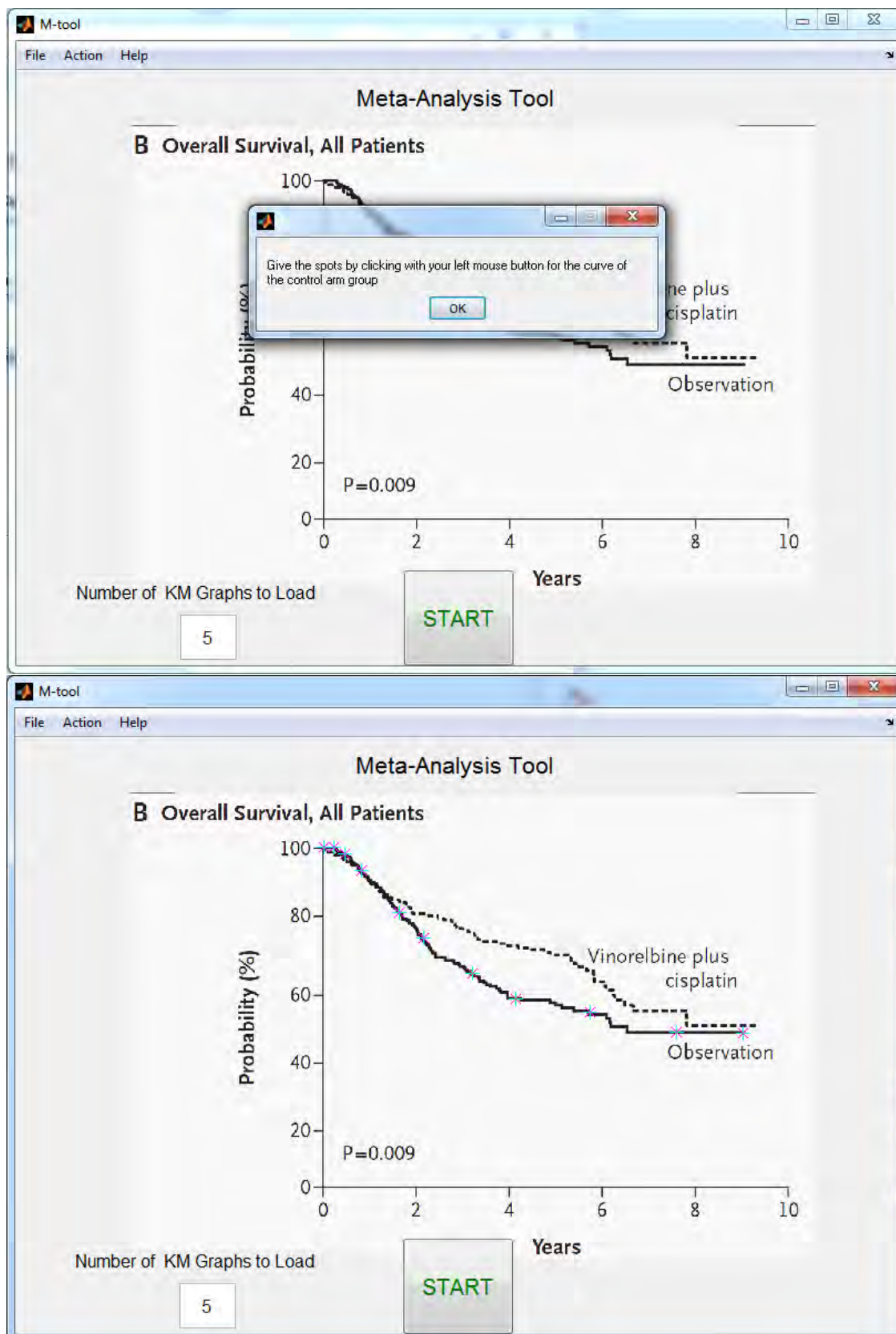


Εικόνα 19 Πατώντας την επιλογή φόρτωσης αρχείου εικόνας και έπειτα την επιλογή φόρτωσης αρχείου txt που περιέχει τα χρονικά σημεία ανοίγει ένα παράθυρο που σου εμφανίζει τα αρχεία αυτά που θα τεθούν προς επεξεργασία.

Αφού έχει ολοκληρωθεί αυτό το βήμα έχει γίνει σωστά η επιλογή αρχείου εικόνας και αρχείου κειμένου των χρονικών σημείων για την πραγματοποίηση της μετα-ανάλυσης. Η μετα-ανάλυση μπορεί να εφαρμοστεί σε 2 ή περισσότερες μελέτες άρα είναι απαραίτητη η γνώση του πλήθους των μελετών που θα συμμετέχουν. Αυτό επιτυγχάνεται με ένα στοιχείο στο γραφικό περιβάλλον της εφαρμογής τύπου text box στο οποίο συμπληρώνουμε το πλήθος των εικόνων που θα συμμετέχουν στην μετα-ανάλυση. Στην εικόνα 19 φαίνεται ο τρόπος χρήσης του συγκεκριμένου κουτιού. Η εφαρμογή τίθεται σε λειτουργία την εφαρμογή μας με την επιλογή Start όπως φαίνεται στην εικόνα 21.

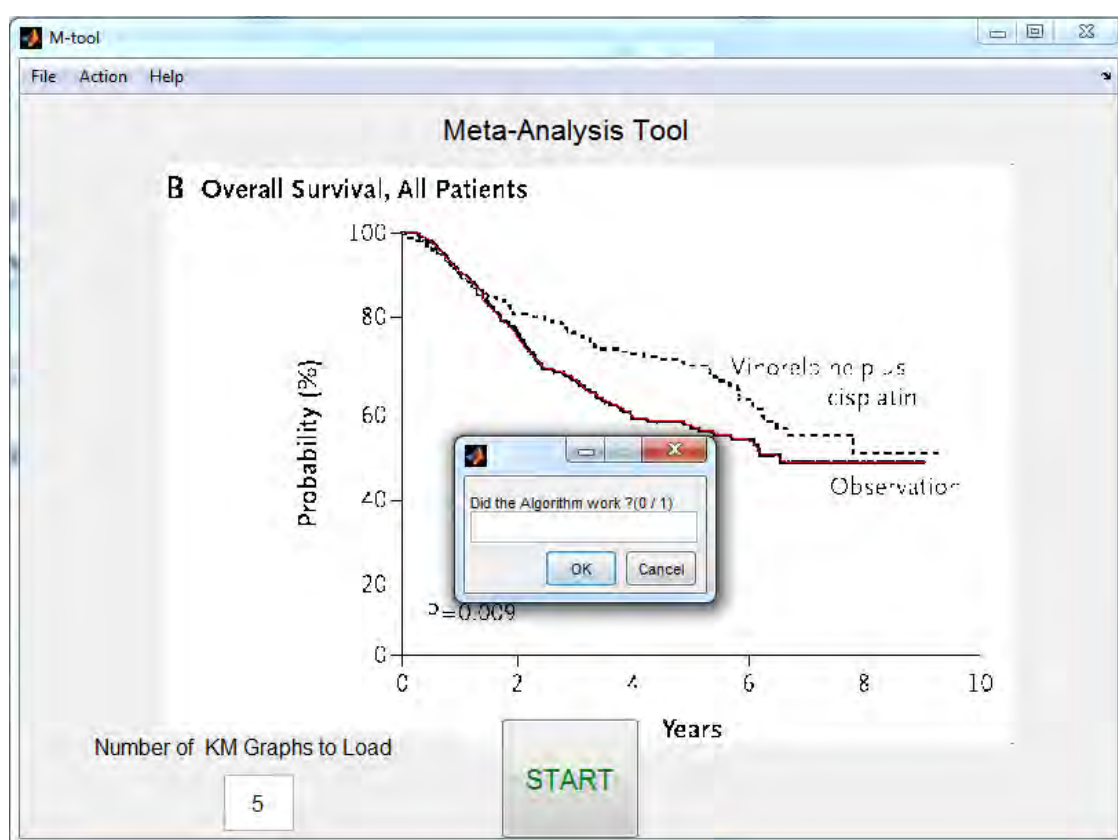


Εικόνα 20 Συμπλήρωση του πλήθους των καμπυλών και έναρξη λειτουργίας της εφαρμογής πατώντας το κουμπί start.



Εικόνα 21 Έναρξη λειτουργίας του αλγορίθμου μετά το πάτημα της ένδειξης ok για να μπορέσει ο χρήστης να δώσει με το ποντίκι τα σημεία που θα διευκολύνουν τον εντοπισμό της καμπύλης ελέγχου.

Στο γραφικό περιβάλλον της εφαρμογής θα εμφανιστεί ένα παράθυρο με την εικόνα που θέσαμε προς επεξεργασία στην οποία θα κάνουμε κάποια κλικ πάνω στην καμπύλη που μας ενδιαφέρει έπειτα από προτροπή της εφαρμογής στο χρήστη και το πάτημα της ένδειξης «OK» για την έναρξη της εφαρμογής όπως φαίνεται στην εικόνα 22. Σκοπός αυτής της διαδικασίας είναι η καθοδήγηση του αλγόριθμου εύρεσης μονοπατιού. Το πλήθος των οδηγιών σημείων που πρέπει να εισάγει ο χρήστης διαφέρει ανάλογα με το συνολικό αριθμό των καμπυλών και την διαδρομή που ακολουθούν. Σε αντίθεση με τα σημεία έναρξης και λήξης που είναι τα ίδια για όλες τις καμπύλες, δηλαδή ισχύει ότι το πρώτο σημείο εισάγεται από το χρήστη στην αρχή της καμπύλης και το τελευταίο στο τέλος αυτής. Σημαντικό είναι το γεγονός ότι όλα τα σημεία θα δίνονται από τον χρήστη με αριστερό κλικ εκτός από το τελευταίο. Ο αλγόριθμος θα εκτελείται μετά την επιλογή του τελευταίου σημείου με δεξί κλικ που ουσιαστικά σηματοδοτεί την λήξη επιλογής σημείων και άρα την έναρξη της εφαρμογής του αλγορίθμου του Dijkstra στην συγκεκριμένη καμπύλη όπως φαίνεται στην εικόνα 20. Αφού τελειώσει η εκτέλεση του συγκεκριμένου αλγορίθμου εμφανίζεται το αποτέλεσμα της διαδικασίας με κόκκινη γραμμή πάνω στην καμπύλη που εφαρμόστηκε ο αλγόριθμος του Dijkstra όπως φαίνεται στην εικόνα 23.

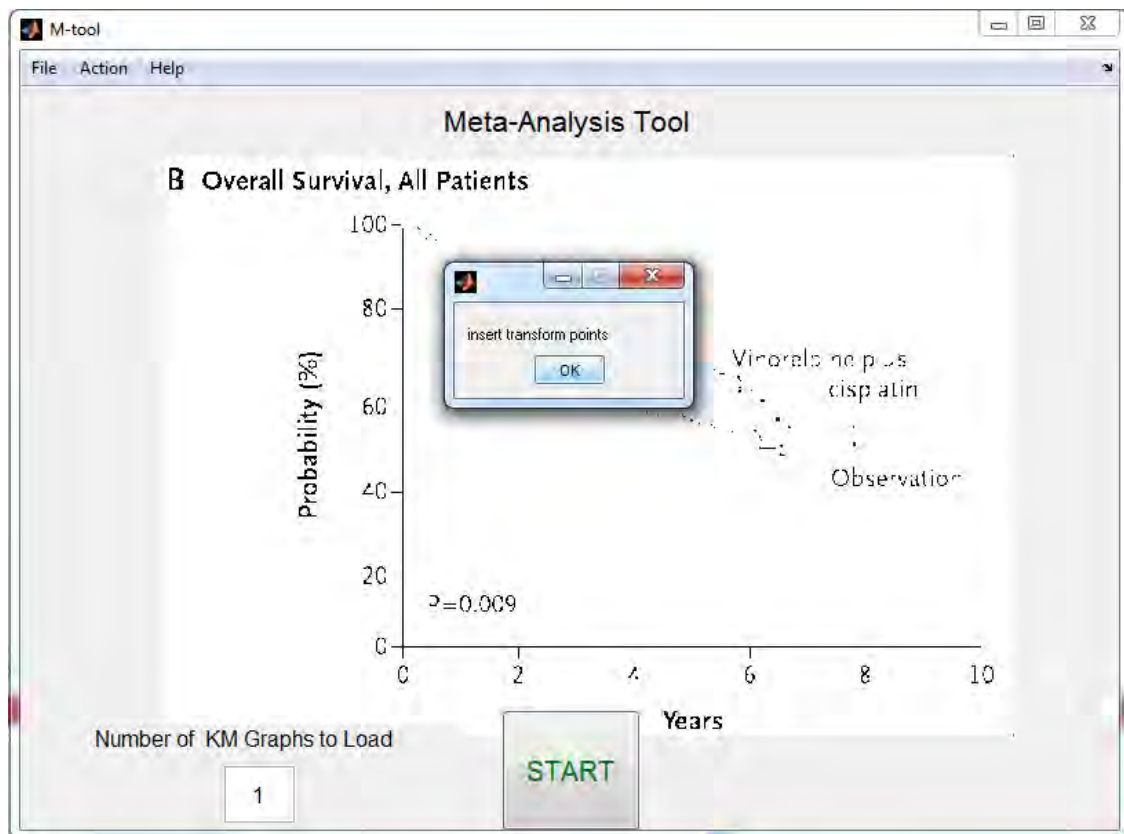


Εικόνα 22 Έλεγχος ορθής λειτουργίας του αλγορίθμου.



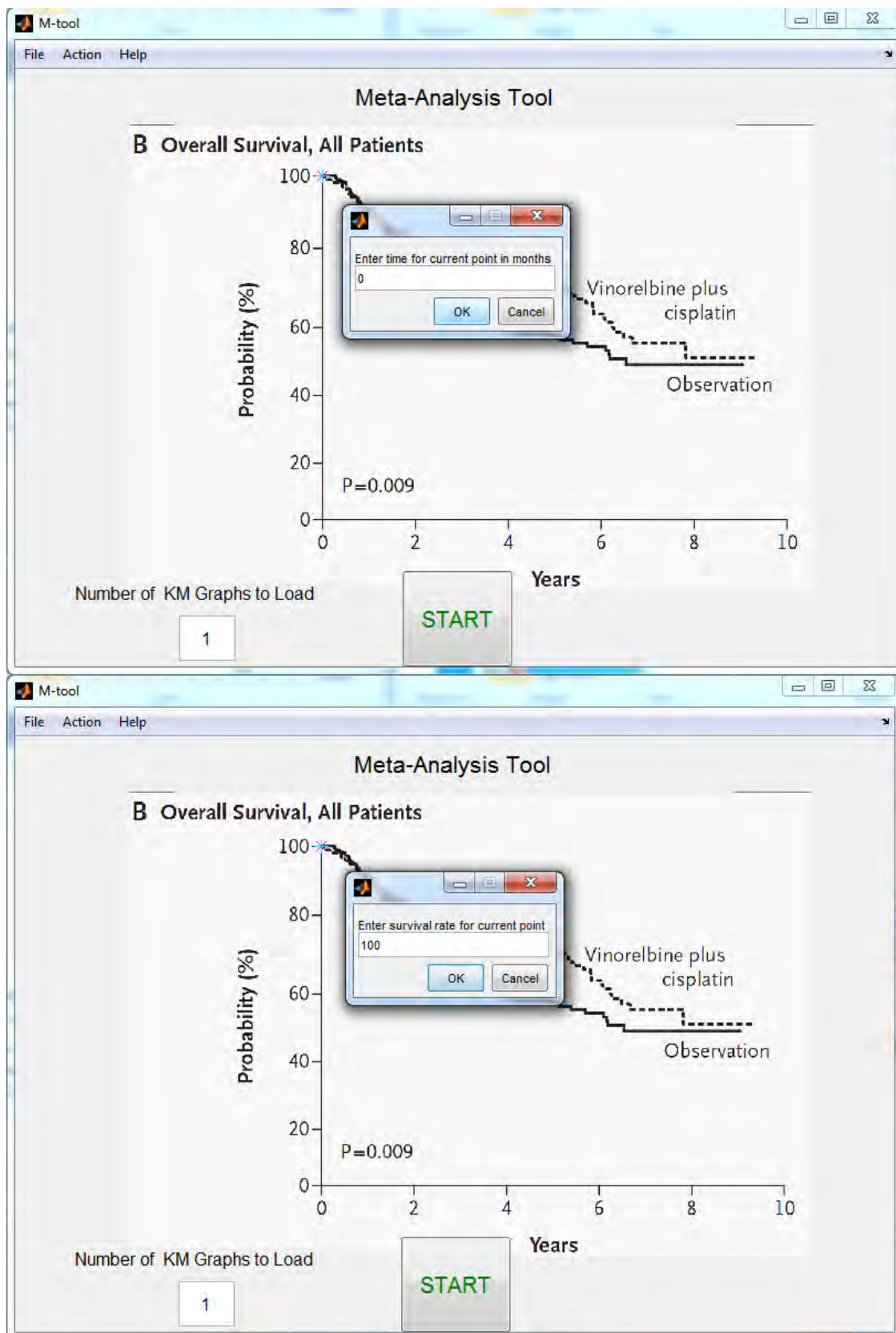
Υπάρχουν περιπτώσεις που ο αλγόριθμος του Dijkstra δεν ακολουθεί την καμπύλη ενδιαφέροντος σωστά σε όλο της το μήκος και η διαδικασία πρέπει να επαναληφθεί. Αυτό γίνεται αντιληπτό με την εμφάνιση του αποτελέσματος της διαδικασίας εύρεσης μονοπατιού και ο χρήστης μπορεί να ζητήσει την επανάληψη της διαδικασίας εύρεσης της καμπύλης δίνοντας τα στοιχεία από την αρχή. Η διαδικασία επαναλαμβάνεται και για τις υπόλοιπες καμπύλες της εικόνας που μας ενδιαφέρουν.

Μετά το πέρας των παραπάνω διαδικασιών εμφανίζεται στην εφαρμογή μήνυμα για να προτρέψει το χρήστη να εισάγει με το ποντίκι τουλάχιστον τρία σημεία πάνω στους άξονες του γραφήματος για την δημιουργία του πίνακα μετασχηματισμού (βλ. Παρ. 2.3.6), όπως φαίνεται στην εικόνα 24. Για κάθε σημείο δίνονται οι πραγματικές τιμές των συντεταγμένων αυτού και η διαδικασία επαναλαμβάνεται για τρεις τουλάχιστον φορές ή παραπάνω για μεγαλύτερη ακρίβεια όπως φαίνεται στην εικόνα 25.



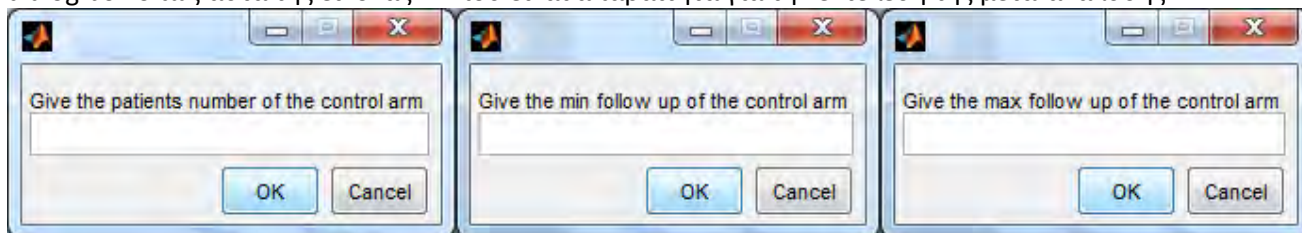
Εικόνα 23 Έναρξη διαδικασίας δημιουργίας πίνακα μετασχηματισμού.





Εικόνα 24 Παράδειγμα εισαγωγής χρόνο (σε μήνες) και ποσοστού επιβίωσης για ένα από τα τρία τουλάχιστον σημεία που πρέπει να δώσει ο χρήστης για την δημιουργία του πίνακα μετασχηματισμού

Αφού έχουν ολοκληρωθεί τα προηγούμενα βήματα εμφανίζονται στην οθόνη οι καμπύλες που προέκυψαν με την εκτέλεση της εφαρμογής μας και ζητείται από τον χρήστη να δώσει κάποια δεδομένα μέσω dialog box όπως αυτά της εικόνας 24 που είναι απαραίτητα για την εκτέλεση της μετα-ανάλυσης.



Εικόνα 25 Dialog boxes που συμπληρώνονται από τον χρήστη για την έναρξη της μετα-ανάλυσης

## 4 Αποτελέσματα

### 4.1 Εισαγωγή

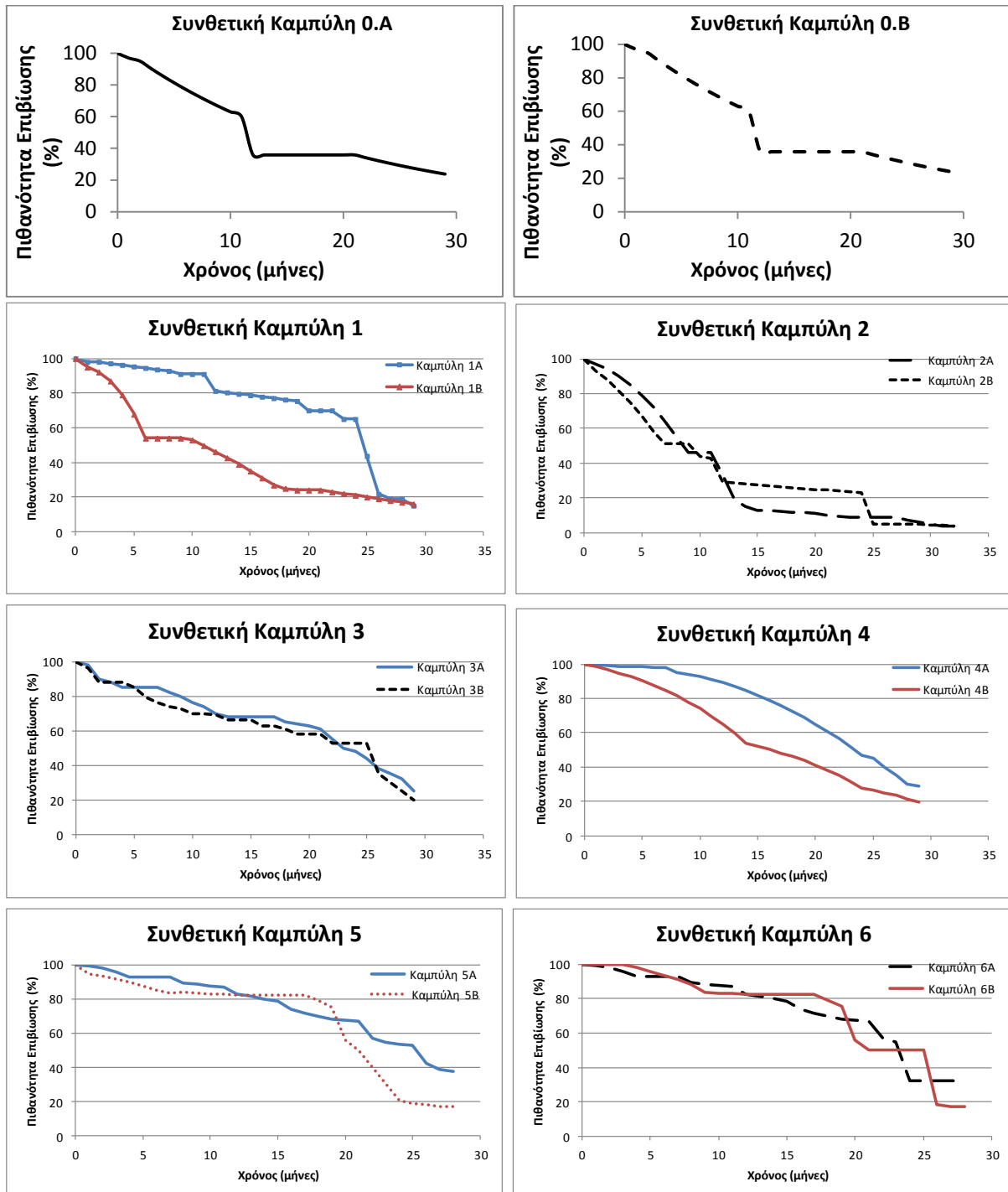
Πριν την ολοκλήρωση του κώδικα και αφού είχαν ήδη πραγματοποιηθεί αρκετές δοκιμές προέκυψε η ανάγκη να ελέγξουμε την αποτελεσματικότητα του τελικού εργαλείου μέσω συνθετικών καμπυλών που δημιουργήθηκαν με αριθμούς που εμείς οι ίδιοι ορίσαμε. Ο έλεγχος αυτός είχε σκοπό την δυνατότητα χρήσης της εφαρμογής σε καμπύλες Kaplan Meier από διάφορες δημοσιεύσεις χωρίς να υπάρχει αμφιβολία για την εγκυρότητα των αποτελεσμάτων. Με τον έλεγχο μέσω των συνθετικών καμπυλών μας δίνεται η δυνατότητα να βελτιώσουμε την ακρίβεια του αλγορίθμου και να μειώσουμε στο ελάχιστο την απόκλιση από τις αρχικές τιμές της καμπύλης.

Με την βοήθεια του εργαλείου Microsoft Excel δημιουργήσαμε ένα σύνολο καμπυλών με διαφορετικά χαρακτηριστικά αναπαράστασης καθώς παίζουν σημαντικό ρόλο στον τρόπο λειτουργίας της εφαρμογής. Έτσι οι καμπύλες που φτιάξαμε ήταν μαύρες ή χρωματιστές, είτε διασταυρώνονται είτε είναι μακριά η μία από την άλλη, με τελείες παύλες ή άλλα σύμβολα και αρκετές ακόμα λεπτομέρειες για να καλύψουμε όλο το φάσμα των καμπυλών Kaplan Meier που μπορεί να υπάρχουν στις δημοσιεύσεις. Η αποθήκευση των καμπυλών γίνεται σε οποιοδήποτε τύπου εικόνας bitmap και χρησιμοποιείται σαν δεδομένο εισόδου στο πρόγραμμα που δημιουργήσαμε σε ψευδογλώσσα του λογισμικού Matlab.

Οι συνθετικές καμπύλες έχουν επίσης συγκεκριμένα τεχνικά χαρακτηριστικά που ανταποκρίνονται σε αυτά των καμπυλών Kaplan Meier, ώστε να έχουμε την δυνατότητα να ελέγξουμε την ακρίβεια των αποτελεσμάτων, την ορθή λειτουργία του αλγορίθμου και το χρόνο εκτέλεσης του σε πραγματικά δεδομένα. Τα χαρακτηριστικά αυτά είναι, η μονοτονία των καμπυλών, η μέγιστη τιμή που αντιστοιχεί στο 100, τα χρονικά διαστήματα που είναι συνήθως ίσα για συντομία στον υπολογισμό τους και ο τρόπος αναπαράστασης τους. Το 100 αντιπροσωπεύει το ποσοστό επιβίωσης κατά την έναρξη της μελέτης, τα χρονικά διαστήματα ξεκινάνε με 0 που εκφράζει ουσιαστικά την πρώτη μέρα της μελέτης και η καμπύλη είναι φθίνουσα γιατί με το πέρασμα του χρόνου μειώνεται το ποσοστό των ασθενών που είναι ζωντανοί και συμμετέχουν στην μελέτη.

Κατά την αναζήτηση στο διαδίκτυο για πληροφορίες που έχουν σχέση με τις μεθόδους πραγματοποίησης μιας μετα-ανάλυσης βρέθηκε σε δημοσίευση εργασίας αναφορά σε εφαρμογή που ακολουθούσε περίπου την ίδια τεχνική με την εφαρμογή της παρούσας εργασίας. Το εργαλείο αυτό είναι ουσιαστικά μια εφαρμογή που ονομάζεται *Digitize it* [27] και λειτουργεί λαμβάνοντας σαν πληροφορία εισόδου μια εικόνα με καμπύλες Kaplan Meier, εντοπίζοντας τα σημεία που σχηματίζουν την καμπύλη και επιστρέφοντας τις πραγματικές τιμές της. Η διαδικασία εντοπισμού της καμπύλης μπορούσε να γίνει με πολλαπλά κλικ πάνω στην καμπύλη, με ένα κλικ πάνω στην καμπύλη, και σε περίπτωση που ήταν σχηματισμένη με σύμβολα γινόταν αναγνώριση αυτών. Επίσης είτε πριν είτε μετά τον εντοπισμό της καμπύλης έπρεπε να δώσεις 4 τιμές από τους άξονες που απεικονίζονταν οι καμπύλες για να παράγει τα τελικά αποτελέσματα. Υπήρχαν μερικά κοινά σημεία με την εφαρμογή που παρουσιάζεται στην παρούσα εργασία και έτσι πραγματοποιήθηκαν κάποιες δοκιμές με συνθετικές καμπύλες και στις δύο εφαρμογές για να ελέγξουμε την εγκυρότητα των αποτελεσμάτων από την εκάστοτε εφαρμογή και να τις συγκρίνουμε μεταξύ τους.

## 4.2 Συνθετικές Καμπύλες



Εικόνα 26 Συνθετικές καμπύλες που χρησιμοποιήθηκαν με τα συγκεκριμένα χαρακτηριστικά και τρόπο απεικόνισης για τον έλεγχο της εφαρμογής της συγκεκριμένης εργασίας.

Οι πρώτες δοκιμές ξεκίνησαν με σκοπό να ελέγξουμε την συμπεριφορά των εφαρμογών σε απλές καμπύλες. Έτσι δημιουργήσαμε γραφήματα με μια καμπύλη που στην μια περίπτωση απεικονίζεται με συνεχόμενη γραμμή και στην άλλη με διακεκομμένες γραμμές και στην συνέχεια πολύ πιο πολύπλοκα όπως φαίνεται στην εικόνα 27. Τα χρονικά διαστήματα που αντιστοιχούν σε συγκεκριμένες στιγμές και χρησιμοποιήθηκαν για την δημιουργία των συνθετικών καμπυλών κυμαίνονται από 29 μέχρι και 35. Με βάση τον πλήθος των χρονικών διαστημάτων και των χρονικών στιγμών που τα δημιουργούν αναζητάμε να βρούμε τις πιθανότητες επιβίωσης που αντιστοιχούν σε αυτά με την χρήση δυο διαφορετικών εφαρμογών. Το Digitize it που είναι η μια εφαρμογή, χρειάστηκε τόσα κλικ όσα είναι και τα χρονικά διαστήματα της καμπύλης με πολύ μεγάλη προσοχή και ακρίβεια για τον εντοπισμό των τιμών που μας ενδιέφεραν. Αυτό βέβαια το γεγονός σημαίνει ότι χρειάστηκε πολύ περισσότερο χρόνο για την ολοκλήρωση της διαδικασίας σε σχέση με την εφαρμογή της συγκεκριμένης εργασίας. Υπάρχει βέβαια και η δυνατότητα να εντοπιστεί η καμπύλη με ένα κλικ αλλά θα έχουμε πολλά σημεία και θα πρέπει να γίνει αναζήτηση αυτών που θέλουμε με άλλο τρόπο σε αντίθεση με την αναζήτηση στην εφαρμογή που αναπτύσσεται στην συγκεκριμένη εργασία που γίνεται αυτόματα.

Εκτός από τα δύο πρώτα γραφήματα της εικόνας 27 που περιέχουν μια καμπύλη τα επόμενα έχουν από 2 καμπύλες και είναι πιο πολύπλοκα. Οι συνθετικές καμπύλες Α και Β καθώς επίσης και η συνθετική καμπύλη 1 χρησιμοποιήθηκαν για σύγκριση για τις δύο εφαρμογές ενώ οι υπόλοιπες για τον έλεγχο της εφαρμογής της παρούσας εργασίας. Τα αποτελέσματα από την σύγκριση των δυο εφαρμογών φαίνονται στον πίνακα 5, 6 καθώς επίσης και στις εικόνες 28, 29 αντίστοιχα ενώ αυτά από τον έλεγχο μόνο της εφαρμογής της εργασίας παρουσιάζονται στον πίνακα 5 και στο σύνολο εικόνων 30, 31. Τα αποτελέσματα στους πίνακες παρουσιάζουν το σφάλμα που ορίζεται σαν η απόλυτη τιμή της διαφοράς των πιθανοτήτων επιβίωσης που προκύπτουν από την εφαρμογή σε σχέση με τις πραγματικές πιθανοτήτων επιβίωσης των συνθετικών καμπυλών. Η πιθανότητα επιβίωσης είναι διαφορετική για κάθε χρονικό διάστημα έτσι υπολογίζουμε τόσες διαφορές όσες και το πλήθος των χρονικών διαστημάτων. Αφού έχουν πραγματοποιηθεί οι παραπάνω υπολογισμοί για να μην παρατεθούν όλα τα στοιχεία του σφάλματος αναφέρονται αυτά που είναι ικανά για την αποτελεσματική περιγραφή του όπως το μέσο, το μεσαίο και το μέγιστο σφάλμα.

**Πίνακας 5 Συγκριτικό σφάλμα ανάλογα με τα κλικ που εισάγει ο χρήστης για την εφαρμογή Digitize It σε σχέση με την εφαρμογή της παρούσας εργασίας για την συνθετική καμπύλη Α και για την συνθετική καμπύλη Β.**

Συνθετική Καμπύλη 0.A	Πλήθος Κλικ από τον Χρήστη							
	2		5		12		30	
Σφάλμα	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt
Μέσο	0,24	--	0,24	--	0,30	--	0,26	0,11
Μέγιστο	1,6	--	0,93	--	0,95	--	1,48	0,29
Μεσαίο	0,15	--	0,18	--	0,22	--	0,21	0,08
Συνθετική Καμπύλη 0.B	Πλήθος Κλικ από τον Χρήστη							
	2		5		12		30	
Σφάλμα	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt
Μέσο	0,24	--	0,24	--	0,30	--	0,26	0,13
Μέγιστο	1,6	--	0,93	--	0,95	--	1,48	0,29
Μεσαίο	0,15	--	0,18	--	0,22	--	0,21	0,14

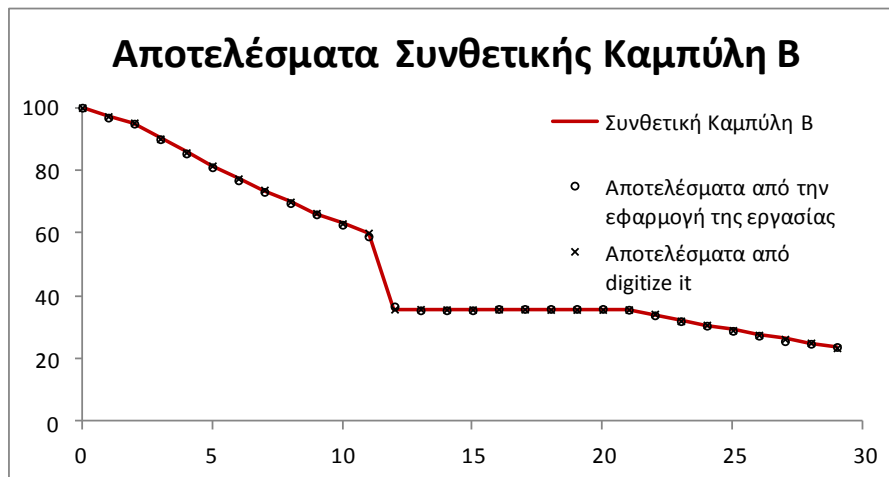
**Πίνακας 6 Συγκριτικό σφάλμα ανάλογα με τα κλικ που εισάγει ο χρήστης για την εφαρμογή Digitize It σε σχέση με την εφαρμογή της παρούσας εργασίας για την συνθετική καμπύλη 1Α και 1Β.**

Συνθετική καμπύλη 1Α	Πλήθος Κλικ από τον Χρήστη							
	2		5		12		30	
Σφάλμα	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt
Μέσο	0,278	--	0,251	--	0,267	--	0,281	0,29
Μέγιστο	1,357	--	1,357	--	1,10	--	1,819	1,11
Μεσαίο	0,146	--	0,143	--	0,130	--	0,142	0,20
Συνθετική Καμπύλη 1Β	Πλήθος Κλικ από τον Χρήστη							
	2		5		12		30	
Σφάλμα	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt	Εφαρμογή	Digitizelt
Μέσο	0,234	--	0,226	--	0,226	--	0,198	0,25
Μέγιστο	0,857	--	0,857	--	0,857	--	0,836	0,7
Μεσαίο	0,190	--	0,166	--	0,166	--	0,118	0,20

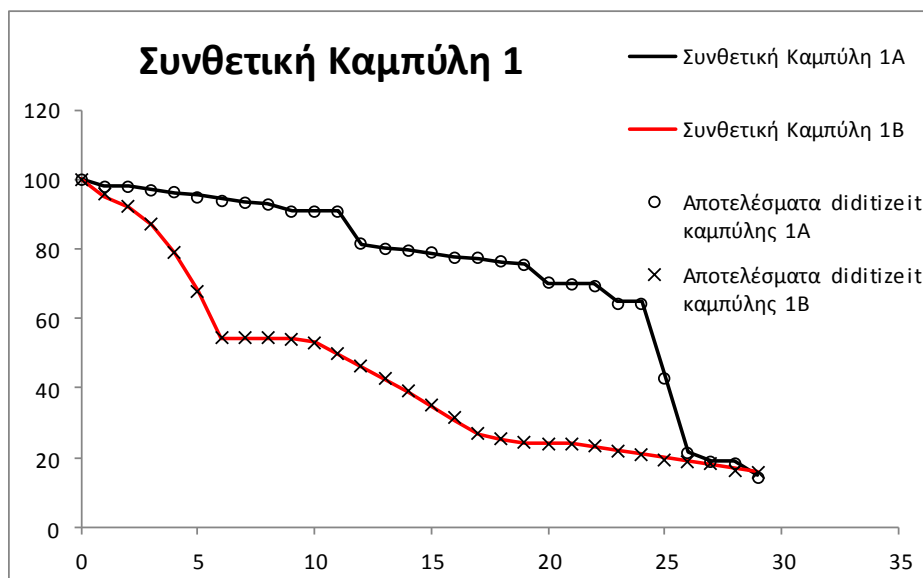
Οι εικόνες απεικονίζουν τις καμπύλες που προκύπτουν από την λειτουργία των εφαρμογών πάνω στις γραφήματα καθώς επίσης και τις συνθετικές καμπύλες έτσι ώστε να υπάρχει και η δυνατότητα οπτικού ελέγχου. Επίσης παρατίθενται γραφήματα που απεικονίζουν την συμπεριφορά του σφάλματος ανάλογα με τα κλικ από τον χρήστη στο σύνολο εικόνων 30, 31. Εύκολα μπορεί να διαπιστωθεί από τα αποτελέσματα στους πίνακες αλλά και από τα γραφήματα σφάλματος ότι ο πλέον κατάλληλος αριθμός κλικ στις εικόνες προς επεξεργασία είναι περίπου στα 5 με 10 και το σφάλμα αρκετά μικρό όπως φαίνεται από τις εικόνες και τους πίνακες που ακολουθούν.

Αφού δοκιμάσαμε και συγκρίναμε τις δύο εφαρμογές καταλήγουμε στο συμπέρασμα ότι το Digitize it είναι μια πολύ καλή εφαρμογή με αποτελέσματα αρκετά αξιόπιστα. Αυτό ισχύει για καμπύλες με λίγα χρονικά διαστήματα που δεν είναι πολύπλοκες και δεν είναι πολλές σε αριθμό. Όταν αυξάνονται αυτές οι παράμετροι είναι δύσκολο να πραγματοποιηθεί μια μετα-ανάλυση καθώς τα δεδομένα θα είναι πάρα πολλά και η διαδικασία για να καταλήξουμε σε αυτά ακόμα πιο δύσκολη και χρονοβόρα.

Έπειτα από όλες αυτές τις δοκιμές το συμπέρασμα που προκύπτει είναι ότι η συγκεκριμένη εφαρμογή μπορεί να χρησιμοποιηθεί και για πραγματικές καμπύλες επιβίωσης (καμπύλες K-M).



Εικόνα 27 Τα αποτελέσματα σύγκρισης της συνθετικής καμπύλης 0.Α και 0.Β με την εφαρμογή της εργασίας και την εφαρμογή digitize it.

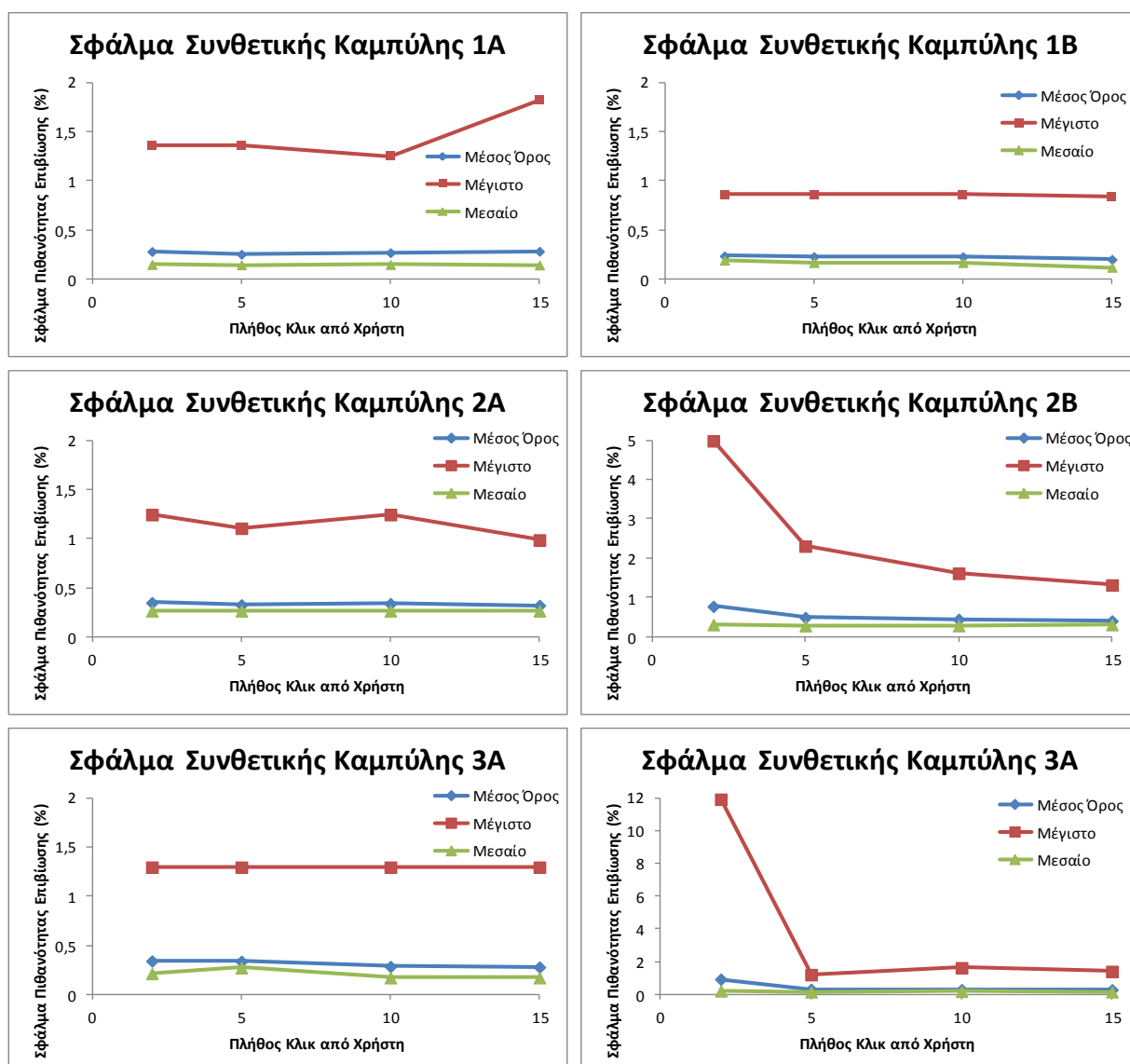


Εικόνα 28 Ένα γράφημα που απεικονίζει τα αποτελέσματα της σύγκρισης της εφαρμογής της εργασίας και της εφαρμογής digitize it για την συνθετική καμπύλη 1.

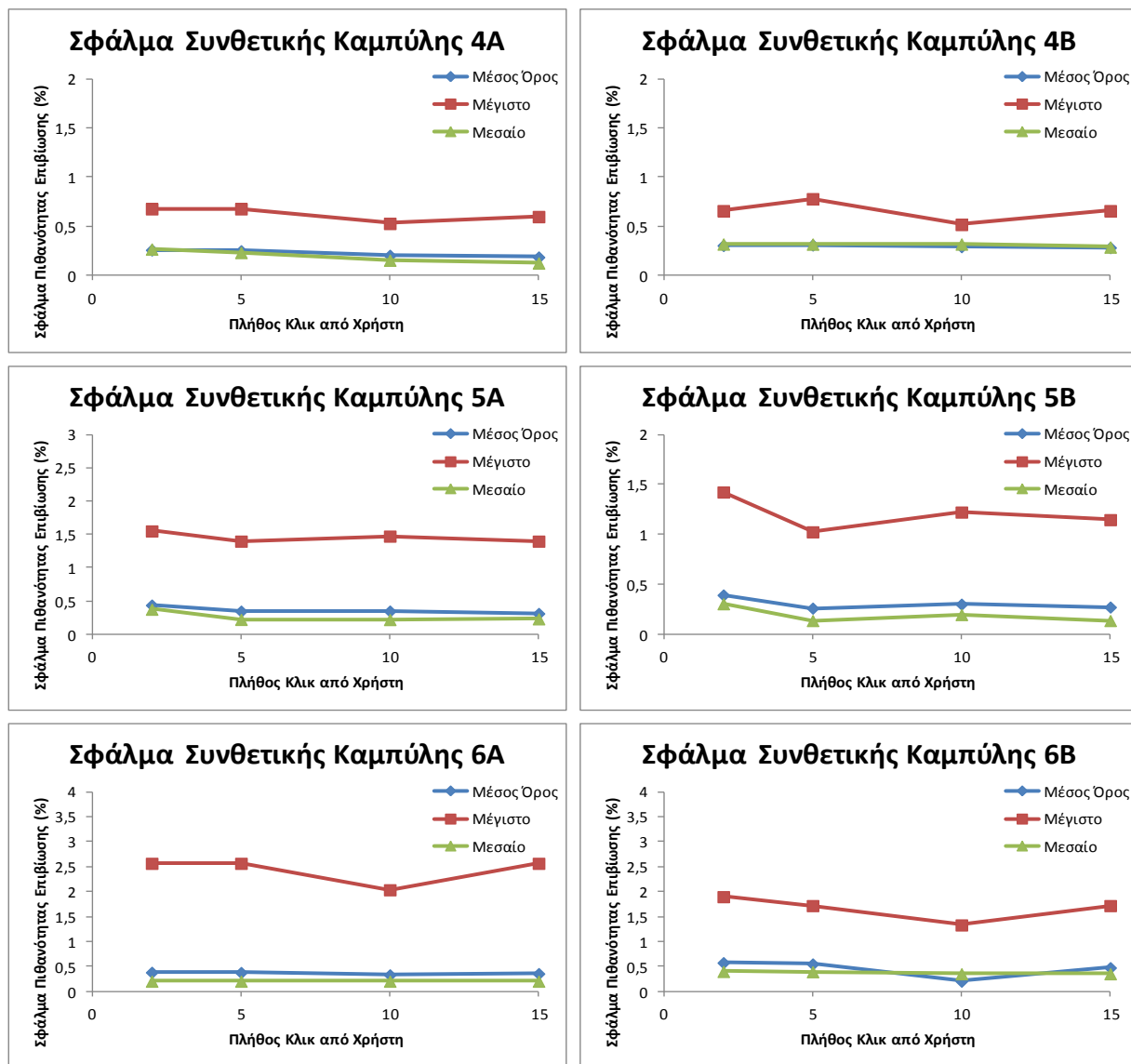
Πίνακας 7 Σφάλμα αυτόματου υπολογισμού πιθανοτήτων συνθετικής καμπύλης του προτεινόμενου αλγόριθμου σαν συνάρτηση του αριθμού των σημείων που εισάγει ο χρήστης.

Συνθετική Καμπύλη	Σφάλμα Υπολογισμού	Πλήθος Κλικ από τον Χρήστη			
		2	5	10	15
1A	Μέσο	0,278	0,251	0,267	0,281
	Μέγιστο	1,357	1,357	1,25	1,819
	Μεσαίο	0,146	0,143	0,146	0,142
1B	Μέσο	0,234	0,226	0,226	0,198
	Μέγιστο	0,857	0,857	0,857	0,836
	Μεσαίο	0,190	0,166	0,166	0,118
2A	Μέσο	0,348	0,325	0,338	0,320
	Μέγιστο	1,243	1,105	1,243	0,986
	Μεσαίο	0,264	0,264	0,264	0,264
2B	Μέσο	0,774	0,490	0,445	0,405
	Μέγιστο	4,985	2,307	1,614	1,319
	Μεσαίο	0,311	0,283	0,285	0,311
3A	Μέσο	0,341	0,337	0,285	0,279
	Μέγιστο	1,295	1,295	1,295	1,295
	Μεσαίο	0,213	0,271	0,171	0,171
3B	Μέσο	0,934	0,314	0,296	0,290
	Μέγιστο	11,915	1,209	1,628	1,403
	Μεσαίο	0,221	0,158	0,192	0,163
4A	Μέσο	0,256	0,246	0,200	0,183
	Μέγιστο	0,675	0,675	0,526	0,595
	Μεσαίο	0,264	0,230	0,151	0,123
4B	Μέσο	0,302	0,307	0,290	0,280
	Μέγιστο	0,655	0,777	0,515	0,655
	Μεσαίο	0,316	0,315	0,316	0,285
5A	Μέσο	0,438	0,347	0,348	0,312
	Μέγιστο	1,551	1,398	1,473	1,398
	Μεσαίο	0,375	0,222	0,220	0,231
5B	Μέσο	0,393	0,261	0,301	0,272
	Μέγιστο	1,424	1,025	1,223	1,147
	Μεσαίο	0,305	0,137	0,197	0,137
6A	Μέσο	0,385	0,381	0,334	0,359
	Μέγιστο	2,567	2,567	2,037	2,567
	Μεσαίο	0,206	0,206	0,206	0,206
6B	Μέσο	0,578	0,552	0,206	0,478
	Μέγιστο	1,899	1,720	1,33	1,720
	Μεσαίο	0,401	0,396	0,355	0,355

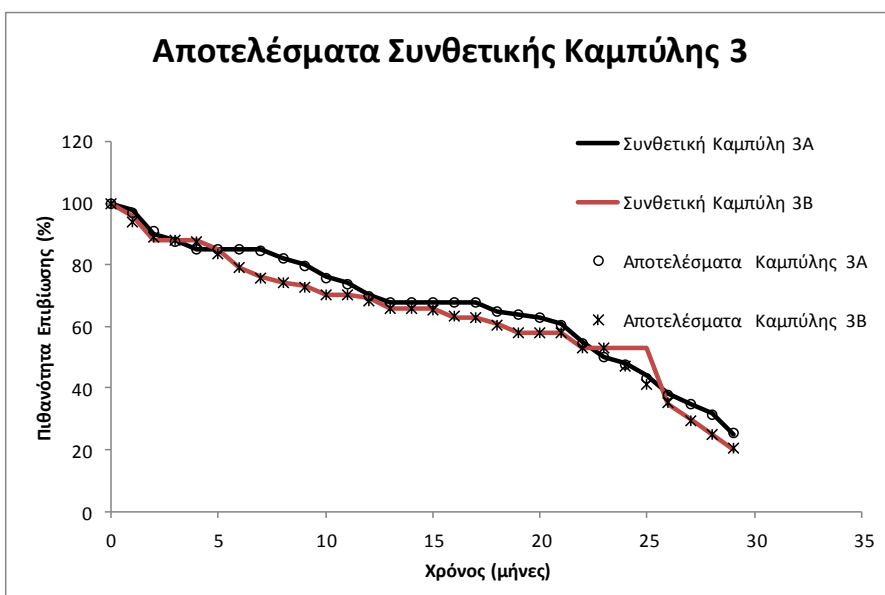
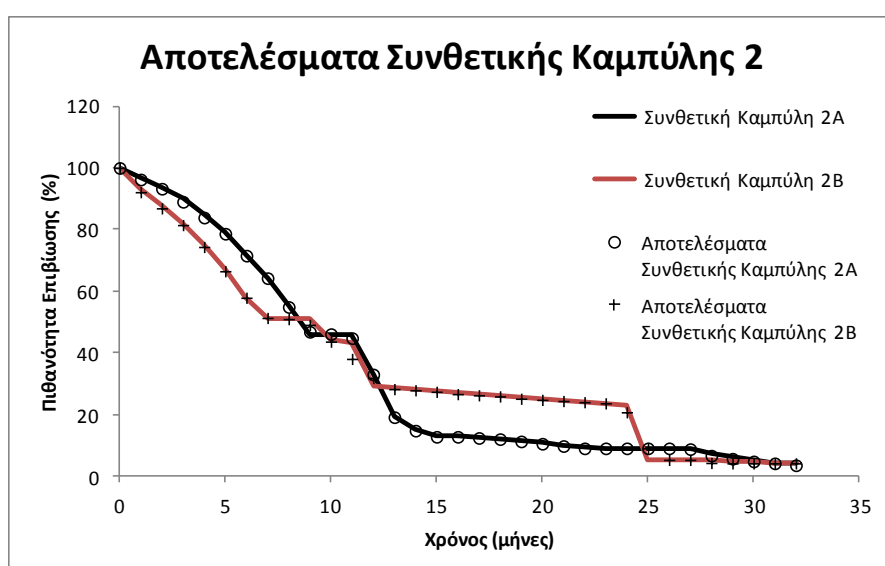
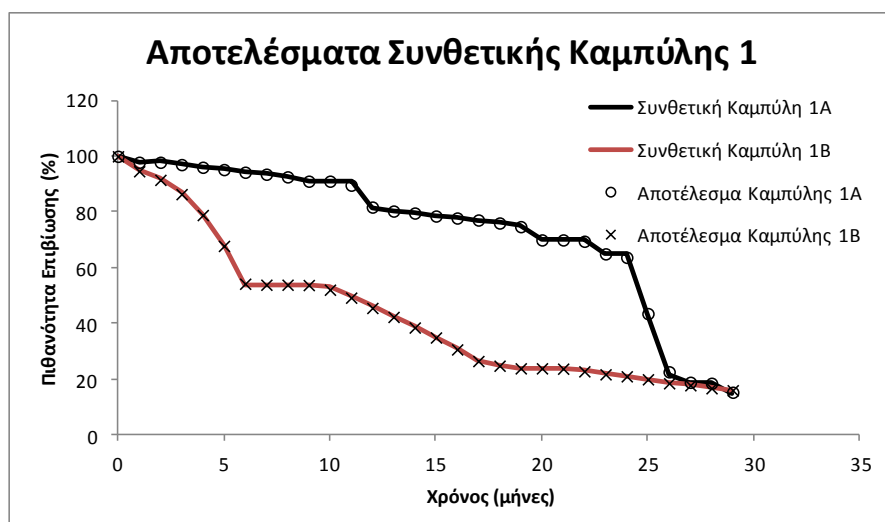




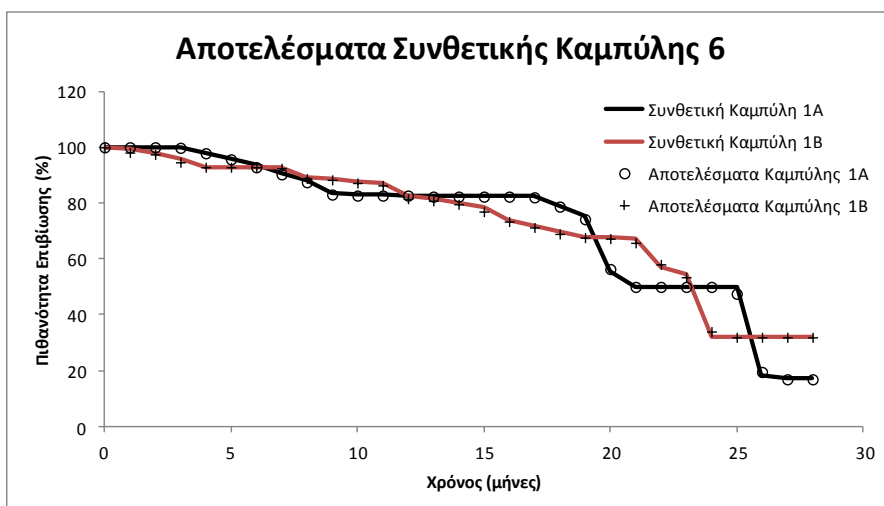
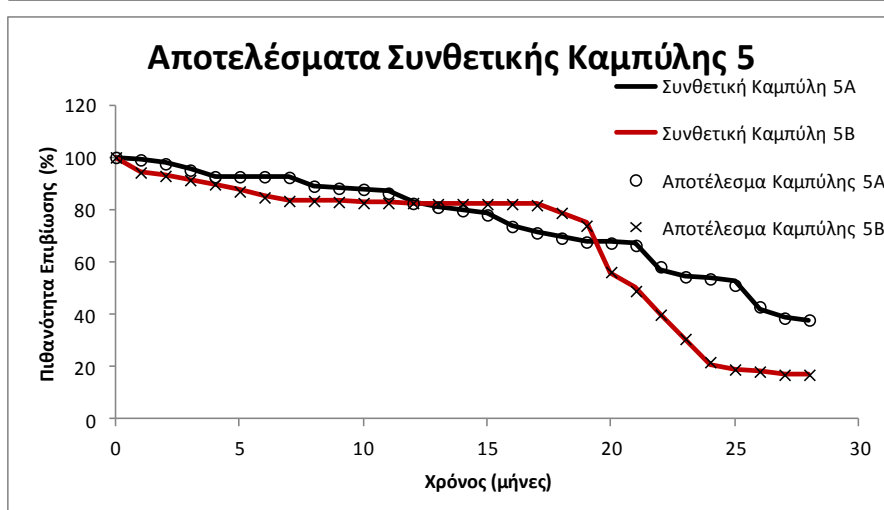
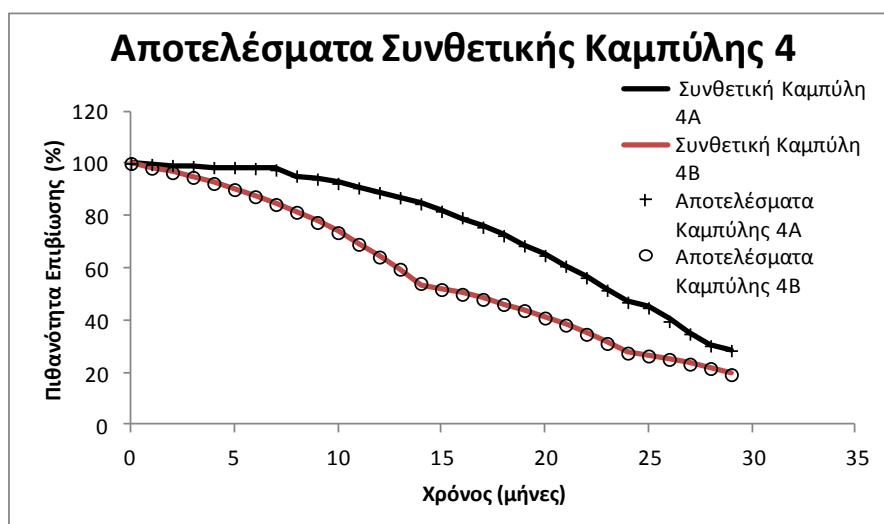
Εικόνα 29 Αποτελέσματα σφάλματος από τις συνθετικές καμπύλες με την χρήση της εφαρμογής της παρούσας εργασίας.



Εικόνα 30 Αποτελέσματα σφάλματος από τις συνθετικές καμπύλες με την χρήση της εφαρμογής της παρούσας εργασίας.



Εικόνα 31 Συνθετικές καμπύλες μαζί με τα αποτελέσματα από την εφαρμογή της παρούσας εργασίας με τον χρήστη να έχει δώσει 8 κλικ με το ποντίκι.



Εικόνα 32 Συνθετικές καμπύλες μαζί με τα αποτελέσματα από την εφαρμογή της παρούσας εργασίας με τον χρήστη να έχει δώσει 8 κλικ με το ποντίκι

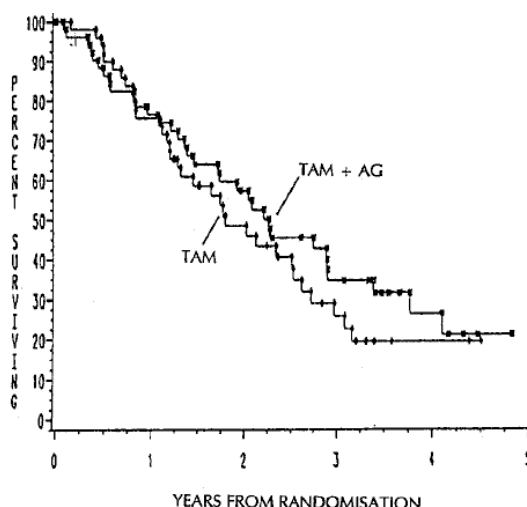
### 4.3 Αποτελέσματα από πραγματικές καμπύλες

Στο συγκεκριμένο υποκεφάλαιο θα γίνει μια παρουσίαση για τον τρόπο που λειτουργεί η εφαρμογή της συγκεκριμένης εργασίας με πραγματικά δεδομένα και κατά πόσο αυτά τα δεδομένα προσεγγίζουν τα πραγματικά αποτελέσματα των μελετών και των μετα-αναλύσεων. Αρχικά θα γίνει μια παρουσίαση ενός πραγματικού παραδείγματος που είναι γνωστά όλα τα στατιστικά μεγέθη που χρειάζονται για την πραγματοποίηση μιας μετα-ανάλυσης τα οποία θα συγκριθούν με τα δεδομένα εξόδου της εφαρμογής. Έπειτα θα παρουσιαστούν τα αποτελέσματα από μεμονωμένες μελέτες που έχουν σαν αντικείμενο τον καρκίνο στους πνεύμονες και τον έλεγχο αποτελεσματικότητας κάποιου νέου φαρμάκου σε σχέση με την παλιά μέθοδο ίασης. Τα αποτελέσματα των οποίων πάλι θα συγκριθούν με αυτά που έχει σαν έξοδο η εφαρμογή που αναπτύχθηκε για την συγκεκριμένη εργασία.

#### 4.3.1 Πραγματικό παράδειγμα Υπολογισμού του Log Hazard Ratio και της διακύμανσης από δημοσιευμένες καμπύλες

Το παράδειγμα που ακολουθεί αναφέρεται σε μια μελέτη εύρεσης καινούργιας θεραπείας για την καταπολέμηση του καρκίνου του μαστού. Η συγκεκριμένη μελέτη έχει ήδη πραγματοποιηθεί και τα αποτελέσματα αυτής έχουν δημοσιοποιηθεί σε εργασία που αναπτύσσει την μέθοδο μετα-ανάλυσης από δημοσιευμένες καμπύλες [16]. Τα δεδομένα της δημοσιευμένης μελέτης χρησιμοποιήθηκαν σαν στοιχεία εισόδου έτσι ώστε να πραγματοποιηθεί έλεγχος και σύγκριση των αποτελεσμάτων που προκύπτουν από την εφαρμογή της παρούσας εργασίας με τα πραγματικά αποτελέσματα από την δημοσιευμένη καμπύλη. Τα στοιχεία αναφέρουν ότι, οι ασθενείς που επιλέχθηκαν τυχαία ήταν στο σύνολο 102 και συγκεκριμένα γυναίκες. Τελικά αυτές που πληρούσαν τις προδιαγραφές της μελέτης ήταν οι 100, οι οποίες χωρίστηκαν σε 2 ομάδες, στην ομάδα ελέγχου και στην ομάδα έρευνας. Στην ομάδα ελέγχου που αποτελούνταν από 49 γυναίκες, χορηγήθηκε το γνωστό για την καταπολέμηση του καρκίνου του μαστού Tamoxifen (TAM) ενώ στην ομάδα έρευνας που αποτελούνταν από 51 γυναίκες, χορηγήθηκε tamoxifen σε συνδυασμό με αμινογλουταμίνη και υδροκορτιζόνη (TAM+ AG).

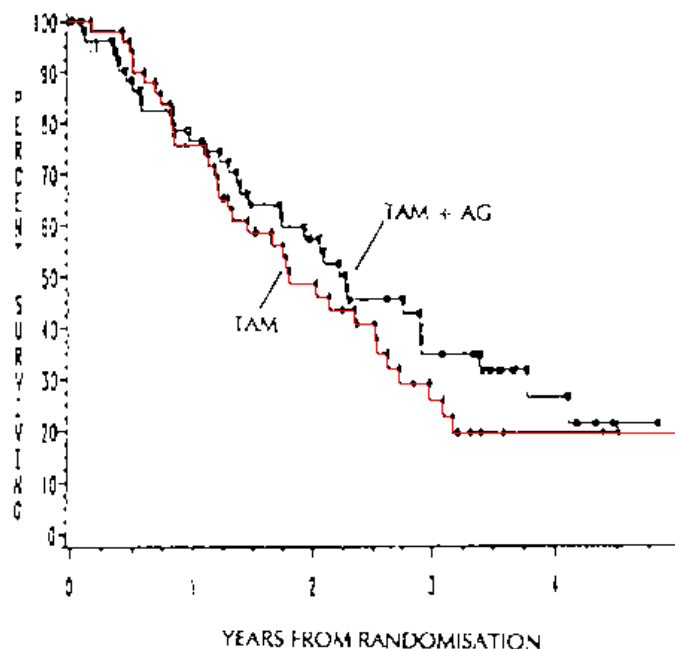
Η εικόνα 34 είναι το αποτέλεσμα της μελέτης αποτυπωμένο σε καμπύλες KM με βάση τα πραγματικά στοιχεία της δημοσιευμένης μελέτης που βρίσκονται στους πίνακες 10,11,12 έπειτα από την εφαρμογή της μεθόδου υπολογισμού του log Hazard ratio και της διασποράς από δημοσιευμένες καμπύλες [16] που περιγράφεται στο κεφάλαιο 2.2. Με δεδομένα την εικόνα 34 και τα χρονικά σημεία που έχουν οριστεί κατά την διάρκεια της μελέτης χρησιμοποιήσαμε την εφαρμογή που αναπτύχθηκε στην παρούσα εργασία. Υπολογίσαμε τα απαιτούμενα στατιστικά μεγέθη για την καμπύλη TAM και την καμπύλη TAM+AG τα οποία παρατίθενται στους πίνακες 8 και 9 αντίστοιχα. Τα δεδομένα των πινάκων 8 και 9 έχουν προκύψει από την λειτουργία εύρεσης μονοπατιού της εφαρμογής (κόκκινη γραμμή) όπως φαίνεται στις εικόνες 35 και 36 αντίστοιχα.



Εικόνα 33 Καμπύλη KM από δημοσιευμένη μελέτη και παράδειγμα [16]

Πίνακας 8 Περιέχει τα αποτελέσματα της μελέτης με βάση την εφαρμογή της συγκεκριμένης εργασίας για τον καρκίνο του μαστού που αντιστοιχούν στην καμπύλη επιβίωσης του πληθυσμού που τους χορηγήθηκε Tamoxifen δηλαδή της εικόνας 34.

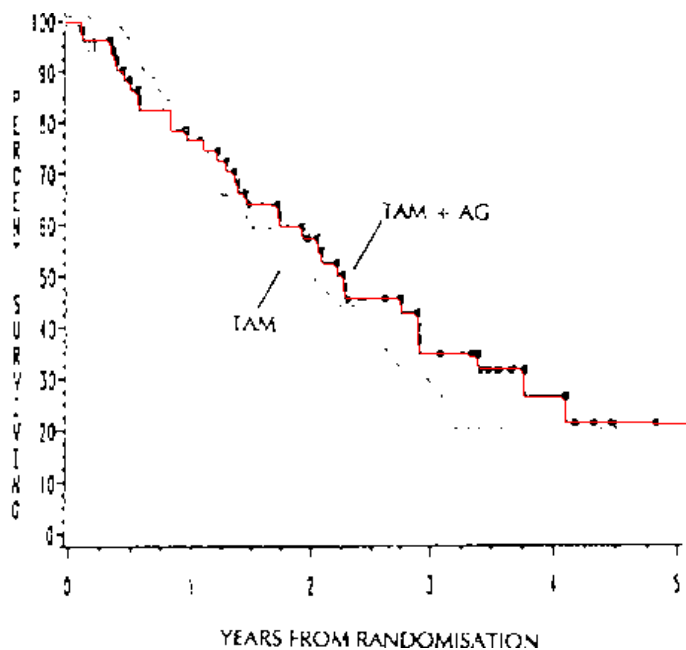
TAM					
Χρονικά διαστήματα σε μήνες	Πιθανότητα επιβίωσης στην αρχή του χρονικού διαστήματος	Ζωντανοί στην αρχή του χρονικού διαστήματος	Censored ασθενείς του χρονικού διαστήματος	Ζωντανοί κατά την διάρκεια του χρονικού διαστήματος	Θάνατοι του χρονικού διαστήματος
$t$	$S_{ci}(t_s)$	$R_{ci}(t_s)$	$C_{ci}(t)$	$R_{ci}(t)$	$D_{ci}(t)$
0-3	100	49	0	49	0,98
3-6	98,00	48,02	0,00	48,02	1,54
6-9	94,86	46,48	0,00	46,48	4,46
9-12	85,77	42,03	0,00	42,03	4,60
12-15	76,39	37,43	0,94	36,50	4,89
15-18	66,16	31,61	0,83	30,78	3,04
18-21	59,62	27,74	0,77	26,97	1,16
21-24	57,06	25,80	0,76	25,04	3,12
24-27	49,95	21,92	0,69	21,24	2,18
27-30	44,83	19,06	0,64	18,43	1,17
30-36	41,97	17,25	1,23	16,02	5,43
36-42	27,76	10,60	0,88	9,71	2,19
42-48	21,49	7,52	0,75	6,77	0,01
48-60	21,45	6,76	1,69	5,07	0,01



Εικόνα 34 Καμπύλη επιβίωσης που απεικονίζει το αποτέλεσμα της λειτουργίας του αλγορίθμου εύρεσης μονοπατιού για τον πληθυσμό ελέγχου (TAM)

Πίνακας 9 Περιέχει τα αποτελέσματα της μελέτης με βάση την εφαρμογή της συγκεκριμένης εργασίας για τον καρκίνο του μαστού που αντιστοιχούν στην καμπύλη επιβίωσης του πληθυσμού που τους χορηγήθηκε Tamoxifen σε συνδυασμό με αμινογλουταμίνη και υδροκορτιζόνη δηλαδή της εικόνας 34.

TAM+AG					
Χρονικά διαστήματα σε μήνες	Πιθανότητα επιβίωσης στην αρχή του χρονικού διαστήματος	Ζωντανοί στην αρχή του χρονικού διαστήματος	Censored ασθενείς του χρονικού διαστήματος	Ζωντανοί κατά την διάρκεια του χρονικού διαστήματος	Θάνατοι του χρονικού διαστήματος
$t$	$S_{ri}(t_s)$	$R_{ri}(t_s)$	$C_{ri}(t)$	$R_{ri}(t)$	$D_{ri}(t)$
0-3	100	51	0,00	51	1,89
3-6	96,30	49,11	0,00	49,11	4,06
6-9	88,34	45,05	0,00	45,05	2,76
9-12	82,93	42,30	0,00	42,30	2,90
12-15	77,24	39,39	0,98	38,41	1,98
15-18	73,26	36,43	0,96	35,47	3,99
18-21	65,01	31,48	0,87	30,60	1,68
21-24	61,45	28,93	0,85	28,08	1,37
24-27	58,46	26,71	0,83	25,87	3,21
27-30	51,21	22,66	0,76	21,91	1,77
30-36	47,08	20,14	1,44	18,70	4,18
36-42	36,55	14,52	1,21	13,31	1,04
42-48	33,68	12,27	1,23	11,04	1,78
48-60	28,26	9,26	2,32	6,95	1,27



Εικόνα 35 Καμπύλη επιβίωσης που απεικονίζει τα αποτελέσματα της λειτουργίας του αλγορίθμου εύρεσης μονοπατιού για τον πληθυσμό υπό έρευνα (TAM+AG)

**Πίνακας 10 Περιέχει τα πραγματικά αποτελέσματα της μελέτης για τον καρκίνο του μαστού που αντιστοιχούν στην καμπύλη επιβίωσης του πληθυσμού που τους χορηγήθηκε Tamoxifen δηλαδή της εικόνας 34.**

TAM					
Χρονικά διαστήματα σε μήνες	Πιθανότητα επιβίωσης στην αρχή του χρονικού διαστήματος	Ζωντανοί στην αρχή του χρονικού διαστήματος	Censored ασθενείς του χρονικού διαστήματος	Ζωντανοί κατά την διάρκεια του χρονικού διαστήματος	Θάνατοι του χρονικού διαστήματος
$t$	$S_{ci}(t_s)$	$R_{ci}(t_s)$	$C_{ci}(t)$	$R_{ci}(t)$	$D_{ci}(t)$
0-3	100	49	0	49	0,98
3-6	98	48,02	0	48,02	3,92
6-9	90	44,1	0	44,1	2,94
9-12	84	41,16	0	41,16	3,92
12-15	76	37,24	0,93	36,31	4,78
15-18	66	31,53	0,83	30,7	3,26
18-21	59	27,44	0,76	26,68	2,26
21-24	54	24,42	0,72	23,7	1,32
24-27	51	22,38	0,7	21,68	2,98
27-30	44	18,7	0,62	18,08	0,82
30-36	42	17,26	1,23	16,03	6,11
36-42	26	9,92	0,83	9,09	2,1
42-48	20	7	0,7	6,3	0
48-60	20	6,3	1,57	4,72	0



**Πίνακας 11** Περιέχει τα πραγματικά αποτελέσματα της μελέτης για τον καρκίνο του μαστού που αντιστοιχούν στην καμπύλη επιβίωσης του πληθυσμού που τους χορηγήθηκε Tamoxifen σε συνδυασμό με αμινογλουταμίνη και υδροκορτιζόνη δηλαδή της εικόνας 34.

TAM+AG					
Χρονικά διαστήματα σε μήνες	Πιθανότητα επιβίωσης στην αρχή του χρονικού διαστήματος	Ζωντανοί στην αρχή του χρονικού διαστήματος	Censored ασθενείς του χρονικού διαστήματος	Ζωντανοί κατά την διάρκεια του χρονικού Διαστήματος	Θάνατοι του χρονικού διαστήματος
$t$	$S_{ri}(t_s)$	$R_{ri}(t_s)$	$C_{ri}(t)$	$R_{ri}(t)$	$D_{ri}(t)$
0-3	100	51	0	51	2,04
3-6	96	48,96	0	48,96	4,08
6-9	88	44,88	0	44,88	3,06
9-12	82	41,82	0	41,82	2,55
12-15	77	39,27	0,98	38,29	1,99
15-18	73	36,3	0,96	35,34	4,36
18-21	64	30,99	0,86	30,13	1,41
21-24	61	28,71	0,84	27,87	1,37
24-27	58	26,5	0,83	25,67	3,1
27-30	51	22,57	0,75	21,82	1,71
30-36	47	20,11	1,44	18,67	4,37
36-42	36	14,3	1,19	13,11	1,09
42-48	33	12,02	1,2	10,82	1,97
48-60	27	8,85	2,21	6,64	1,23

**Πίνακας 12 Παρουσιάζονται τα πραγματικά αποτελέσματα και αυτά που προέκυψαν από την εφαρμογή για την παραπάνω μελέτη με θέμα τον καρκίνο του μαστού.**

Χρονικά διαστήματα σε μήνες	Πραγματικά αποτελέσματα από μελέτη για τον καρκίνο του μαστού				Αποτελέσματα από εφαρμογή για την μελέτη του καρκίνου του μαστού			
	Log hazard ratio για το χρονικό διάστημα $t$ $\ln(HR_i(t))$	Διασπορά του log hazard ratio για το χρονικό διάστημα $\text{Var}[\ln(HR_i(t))]$	Λόγος του hazard ratio προς διασπορά	1 /Var	Log hazard ratio για το χρονικό διάστημα $t$ $\ln(HR_i(t))$	Διασπορά του log hazard ratio για το χρονικό διάστημα $\text{Var}[\ln(HR_i(t))]$	Λόγος του hazard ratio προς διασπορά	1 /Var
0-3	0,69	1,47	0,47	0,68	0,62	1,51	0,41	0,66
3-6	0,02	0,46	0,04	2,18	0,95	0,86	1,11	1,17
6-9	0,02	0,62	0,04	1,61	-0,45	0,54	-0,83	1,84
9-12	-0,45	0,60	-0,74	1,67	-0,47	0,52	-0,90	1,94
12-15	-0,93	0,66	-1,41	1,52	-0,95	0,66	-1,45	1,52
15-18	0,15	0,48	0,32	2,10	0,13	0,52	0,25	1,93
18-21	-0,59	1,08	-0,55	0,93	0,24	1,39	0,17	0,72
21-24	-0,12	1,41	-0,09	0,71	-0,94	0,98	-0,96	1,02
24-27	-0,13	0,57	-0,23	1,74	0,19	0,69	0,28	1,46
27-30	0,55	1,70	0,32	0,59	0,24	1,32	0,18	0,76
30-36	-0,49	0,28	-1,76	3,62	-0,42	0,31	-1,35	3,25
36-42	-1,02	1,21	-0,85	0,83	-1,06	1,24	-0,86	0,81
42-48	13,95	1000000	0,00	0,00	4,72	102,98	0,05	0,01
48-60	13,68	1000000	0,00	0,00	4,14	68,47	0,06	0,01

Στον πίνακα 12 παρουσιάζεται μια συνοπτική σύγκριση των πραγματικών αποτελεσμάτων [16] του παραπάνω παραδείγματος με τα αποτελέσματα της εφαρμογής της εργασίας. Οι στατιστικές μεταβλητές που περιλαμβάνει ο πίνακας είναι οι απαραίτητες για την πραγματοποίηση μετα-ανάλυσης. Είναι φανερό ότι έχουμε διαφορές σε αρκετά δεδομένα στον πίνακα 12, αλλά το Hazard Ratio που είναι το τελικό αποτέλεσμα είναι 0.79 για την δημοσιευμένη μελέτη και ακριβώς το ίδιο και για την εφαρμογή της παρούσας εργασίας. Το γεγονός αυτό είναι ενθαρρυντικό για την αποτελεσματικότητα της εφαρμογής. Η ουσιαστική ερμηνεία του αποτελέσματος είναι ότι η χορήγηση του Tamoxifen σε συνδυασμό με αμινογλουταμίνη και υδροκορτιζόνη έναντι του Tamoxifen βελτιώνει την αντιμετώπιση της ασθένειας και την συνολική επιβίωση των ασθενών.

#### 4.3.2 Παρουσίαση μελετών για καρκίνο στους πνεύμονες και μετα-ανάλυση

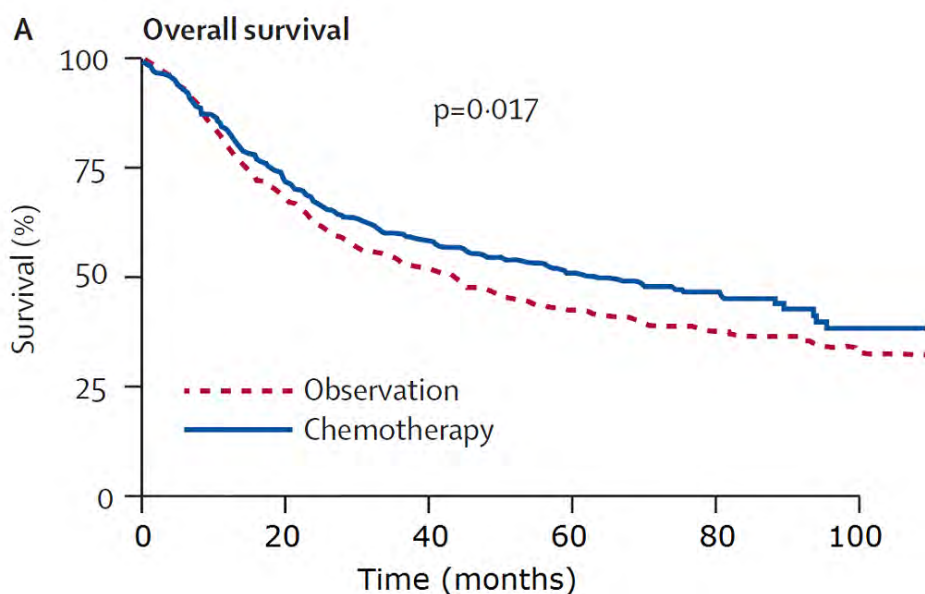
Στο συγκεκριμένο υποκεφάλαιο θα γίνει μια παρουσίαση τυχαιοποιημένων μελετών που διεξήχθησαν σε διάφορα εργαστήρια ανά τον κόσμο με αντικείμενο την θεραπεία του καρκίνου των πνευμόνων. Πιο συγκεκριμένα αρκετές μελέτες έδειξαν ένα σημαντικό όφελος στην συνολική επιβίωση των ασθενών που παρουσίαζαν καρκίνο των πνευμόνων εάν πραγματοποιούνταν χημειοθεραπείες που βασίζονταν στην χορήγηση σισπλατίνης [18] μετά την εγχείρηση αφαίρεσης όγκου του ασθενούς. Οι μελέτες αυτές συγκεντρώθηκαν και αξιοποιήθηκαν για την πραγματοποίηση μιας μετα-ανάλυσης ώστε να υπάρξει η δυνατότητα ενός ασφαλούς συμπεράσματος σε σχέση με την χρήση της σισπλατίνης μετά από την επέμβαση των ασθενών. Αρχικά θα παρουσιαστούν τα αποτελέσματα της εκάστοτε μελέτης όπως προέκυψαν από την εφαρμογή και έπειτα θα γίνει μετα-ανάλυση με βάση αυτά. Στις μελέτες που θα παρουσιαστούν οι ασθενείς χωρίζονται σε δύο βασικούς πληθυσμούς, τον πληθυσμό υπό έρευνα και τον πληθυσμό ελέγχου. Στον πληθυσμό υπό έρευνα εφαρμόζονταν χημειοθεραπείες με βάση την σισπλατίνη ενώ ο πληθυσμός ελέγχου δεν ακολουθούσε κάποια φαρμακευτική αγωγή. Παρακάτω παρουσιάζεται ο πίνακας 13 με όλα τα βασικά στοιχεία από την κάθε μελέτη και έπειτα γίνεται αναφορά στην κάθε μια ξεχωριστά.

**Πίνακας 13 Βασικά στοιχεία των μελετών που θα συμπεριληφθούν στην μετα-ανάλυση.**

Μελέτη	Πλήθος Συμμετεχόντων Στη Μελέτη	Συγγραφέας	Μέση Διάρκεια Παρακολούθησης της Μελέτης (Μήνες)	Χρονική Περίοδος Διεξαγωγής των Μελετών
Adjuvant Navelbine International Trialist Association [28]	840	Jean-Yves Douillard	76	1994-2000
International Adjuvant Lung Cancer Trial [29]	1867	Rodrigo Arriagada	56	1995-2001
Adjuvant Lung Project Italy [30]	1088	Giorgio V. Scagliotti	64,5	1994-1999
Big Lung Trial [31]	307	D. Waller	34,6	1995-2001
JBR10 [32]	482	Timothy Winton	62,4	1994-2001

Πολύ σημαντικοί είναι και οι πίνακες 15, 17, 19, 21, 23 που περιέχουν τα αποτελέσματα του Hazard Ratio της κάθε μελέτης καθώς και τα διαστήματα εμπιστοσύνης από τρεις διαφορετικές πηγές. Οι πηγές αυτές είναι είτε η δημοσίευση της κάθε μελέτης σε επιστημονικά περιοδικά και βιβλιογραφία είτε τα αποτελέσματα από την εφαρμογή της παρούσας εργασίας (με βάση την καμπύλη επιβίωσης της συγκεκριμένης μελέτης) είτε από μια δημοσιευμένη μετα-ανάλυση που χρησιμοποίησε τα στοιχεία της εκάστοτε μελέτης για να καταλήξει στο συνολικό τελικό συμπέρασμα.

Η μελέτη ANITA [28] διήρκησε 10 περίπου χρόνια στο σύνολο και συγκεκριμένα από τον Δεκέμβρη του 1994 μέχρι και τον Αύγουστο του 2004. Σε αυτή συμμετείχαν 840 ασθενείς από τους οποίους οι 407 ανήκαν στον πληθυσμό υπό έρευνα με μια μέση παρακολούθηση 76 μήνες και οι 433 στο πληθυσμό ελέγχου με μια μέση παρακολούθηση 77 μήνες. το Hazard Ratio που προέκυψε είναι της τάξεως του 0,80 (95% CI 0,66-0,96). Δηλαδή η θεραπεία που χρησιμοποιήθηκε στον πληθυσμό υπό έρευνα ήταν αποτελεσματικότερη σε σχέση με την μη χημειοθεραπεία στον πληθυσμό ελέγχου. Το Hazard Ratio όπως και τα περιεχόμενα του πίνακα 14 προέκυψαν από την εφαρμογή που αναπτύχθηκε στην παρούσα εργασία. Τα δεδομένα που χρησιμοποιήθηκαν για να τεθεί σε λειτουργία η εφαρμογή είναι η εικόνα 37 που είναι το αποτέλεσμα της μελέτης με βάση τις καμπύλες K-M, τα χρονικά σημεία που ορίστηκαν από τον χρήστη με τον κανόνα που αναφέρεται στο κεφάλαιο 2 της μεθοδολογίας, βασικά στοιχεία της μελέτης όπως πλήθος ασθενών, ελάχιστη και μέγιστη παρακολούθηση της μελέτης. Για να ελεγχθεί η αποτελεσματικότητα της εφαρμογής ως προς το τελικό αποτέλεσμα γίνεται μια σύγκριση με τα αντίστοιχα αποτελέσματα των δημοσιεύσεων των μελετών και παρουσιάζεται στον πίνακα 15.



Εικόνα 36 Καμπύλη KM τυχαιοποιημένης μελέτης Adjuvant Navelbine International Trialist Association (ANITA) ανοσοενισχυτικής βινoreλβίνης σε συνδυασμό με σισπλατίνη σε ασθενείς με καρκίνο των πνευμόνων.

**14 Αποτελέσματα εφαρμογής από την τυχαιοποιημένη μελέτη Adjuvant Navelbine International Trialist Association [ANITA].**

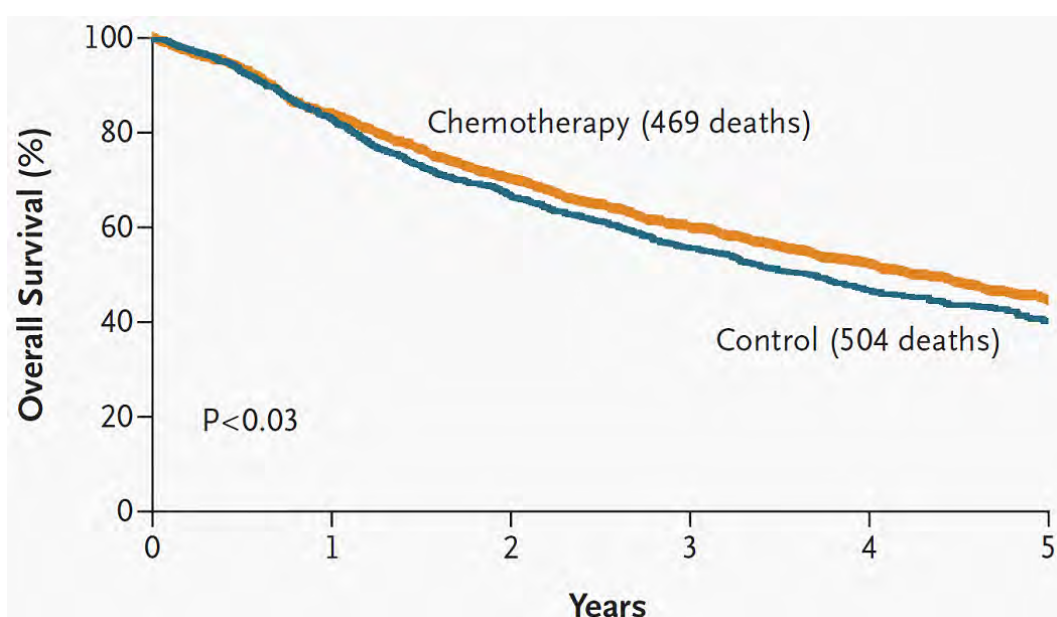
$t_i$	$S_{ci}(t_s)$	$S_{ri}(t_s)$	$R_{ci}(t_s)$	$R_{ri}(t_s)$	$C_{ci}(t)$	$C_{ri}(t)$	$D_{ci}(t)$	$D_{ri}(t)$	$\ln H_i(t)$	$Var_i$
0	100,00	100,00	433,00	407,00	0,00	0,00	3,32	12,53	1,39	0,38
2	99,23	96,92	429,68	394,47	0,00	0,00	30,84	16,44	-0,54	0,09
6	92,11	92,88	398,84	378,03	0,00	0,00	25,42	20,37	-0,17	0,08
9	86,24	87,88	373,42	357,67	0,00	0,00	28,34	19,19	-0,35	0,08
12	79,70	83,16	345,08	338,48	0,00	0,00	24,17	17,62	-0,30	0,09
15	74,11	78,83	312,47	312,41	8,45	8,44	23,52	24,01	0,02	0,08
20	68,54	72,77	276,11	275,58	12,84	12,82	37,59	30,95	-0,19	0,05
28	59,21	64,60	226,89	232,70	11,64	11,93	18,40	14,52	-0,26	0,11
36	54,40	60,57	202,85	212,28	5,63	5,90	9,31	5,71	-0,53	0,27
40	51,91	58,94	179,72	191,81	13,82	14,75	17,28	12,79	-0,37	0,13
50	46,92	55,01	148,90	164,10	13,54	14,92	14,62	10,29	-0,45	0,15
60	42,31	51,56	112,80	129,20	21,49	24,61	7,62	10,54	0,19	0,21

Με τη χρήση των βασικών στοιχείων της μελέτης η εφαρμογή που αναπτύχθηκε είναι σε θέση να έχει πολύ καλά αποτελέσματα όπως φαίνεται στον πίνακα 14. Το Hazard Ratio που προκύπτει είναι ακριβώς το ίδιο με αυτό της μελέτης και το διάστημα εμπιστοσύνης πλησιάζει αρκετά.

**Πίνακας 15 Αποτελέσματα Hazard Ratio από τρεις διαφορετικές πηγές για σύγκριση αποτελεσμάτων στην μελέτη ANITA.**

Μελέτη Adjuvant Navelbine International Trialist Association (ANITA)		
	Hazard Ratio	HR Διαστήματος εμπιστοσύνης 95%
Πηγή: Δημοσιευμένη Μελέτη ANITA [28]	0,80	(0,66-0,96)
Πηγή: Δημοσιευμένη Μετα-ανάλυση [18]	0,82	(0,68-0,98)
Πηγή: Εφαρμογή	0,80	(0,62-0,99)

Τα περιεχόμενα του πίνακα 16 είναι τα αποτελέσματα της μελέτης IALT [29] όπως προέκυψαν από την λειτουργία της εφαρμογής της παρούσας εργασίας. Η συμμετοχή των ασθενών στην συγκεκριμένη μελέτη ξεκίνησε το Φεβρουάριο του 1995 και σταμάτησε να δέχεται άτομα το Δεκέμβρη του 2000. Η λήξη της μελέτης έγινε το 2002. Ο αριθμός των ασθενών που συμμετείχαν τελικά στην μελέτη ανερχόταν στους 1865 από τους οποίους οι 932 ανήκαν στον πληθυσμό υπό έρευνα και οι 935 ανήκαν στον πληθυσμό ελέγχου. Η μέση παρακολούθηση και για τους δύο πληθυσμούς ήταν 56 μήνες. Το αποτέλεσμα της μελέτης αυτής είναι η καμπύλη επιβίωσης K-M της εικόνας 38. Τα παραπάνω βασικά στοιχεία της μελέτης καθώς επίσης η εικόνα 38 και τα χρονικά σημεία (τα στοιχεία της πρώτης στήλης του πίνακα 16) που ορίστηκαν από τον χρήστη είναι τα δεδομένα που χρησιμοποιήθηκαν στην εφαρμογή. Με βάση αυτά τα στοιχεία υπολογίζεται ένα Hazard Ratio της τάξεως του 0,86 (CI 95% 0,76-0,98) από την εφαρμογή. Αυτό σημαίνει ότι η θεραπεία που χρησιμοποιήθηκε στον πληθυσμό υπό έρευνα ήταν αποτελεσματικότερη σε σχέση με τον πληθυσμό ελέγχου που δεν χορηγήθηκε κάποιο φάρμακο. Στον πίνακα 17 πραγματοποιείται μια σύγκριση των αποτελεσμάτων του Hazard Ratio όπως προέκυψε από διαφορετικές πηγές.



Εικόνα 37 Καμπύλη KM από την τυχαιοποιημένη μελέτη International Adjuvant Lung Cancer Trial (IALT).

**Πίνακας 16 Αποτελέσματα εφαρμογής από την τυχαιοποιημένη μελέτη International Adjuvant Lung Cancer Trial (IALT).**

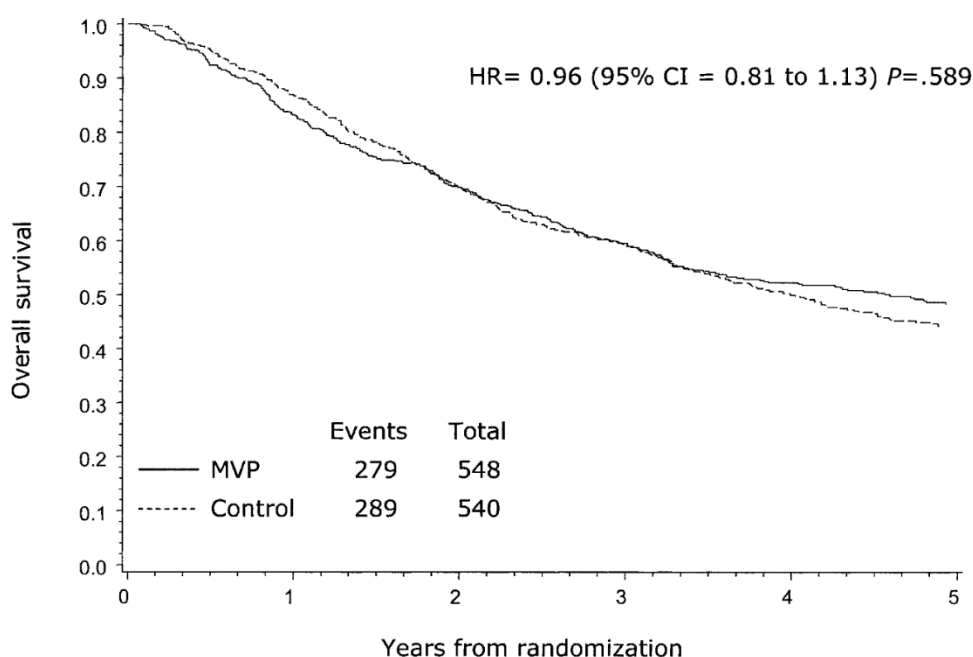
$t_i$	$S_{ci}(t_s)$	$S_{ri}(t_s)$	$R_{ci}(t_s)$	$R_{ri}(t_s)$	$C_{ci}(t)$	$C_{ri}(t)$	$D_{ci}(t)$	$D_{ri}(t)$	$\ln H_i(t)$	$Var_i$
0	100,00	100,00	935,00	932,00	0	0	29,27	33,78	0,15	0,06
4	96,87	96,38	905,73	898,22	0	0	61,10	48,11	-0,23	0,03
8	90,33	91,21	844,63	850,11	0	0	59,43	56,30	-0,06	0,03
12	83,98	85,17	766,05	774,45	19,15	19,36	67,25	48,68	-0,33	0,03
16	76,61	79,82	680,88	707,16	17,92	18,61	46,86	41,23	-0,17	0,04
20	71,33	75,16	616,89	647,93	17,14	18,00	32,41	31,60	-0,07	0,06
24	67,59	71,50	559,43	589,92	25,05	26,41	46,70	44,86	-0,09	0,04
30	61,94	66,06	488,69	519,51	24,03	25,55	44,19	39,51	-0,17	0,04
36	56,34	61,04	421,51	455,17	22,99	24,83	35,13	29,18	-0,26	0,06
42	51,65	57,13	364,09	401,42	22,29	24,58	30,49	26,34	-0,24	0,07
48	47,32	53,38	311,85	350,62	21,76	24,46	20,90	25,70	0,09	0,08
54	44,15	49,46	272,77	304,61	18,18	20,31	17,05	16,49	-0,14	0,11
59	41,39	46,79	0	0	0	0	0	0	0,727	0,245

Τα αποτελέσματα της εφαρμογής που είναι ο πίνακας 16 δεν μπορούν να συγκριθούν με τα πραγματικά καθώς δεν υπάρχουν πληροφορίες σε σχέση με αυτά αλλά μπορεί να ελεγχθεί το τελικό αποτέλεσμα που είναι το Hazard Ratio. Ο πίνακας 17 παρουσιάζει την σύγκριση των αποτελεσμάτων του Hazard Ratio από τρεις διαφορετικές πηγές και όπως είναι προφανές τα αποτελέσματα της μεμονωμένης μελέτης συμφωνούν με αυτά της εφαρμογής και πλησιάζουν πολύ και τα αποτελέσματα των διαστημάτων εμπιστοσύνης.

**Πίνακας 17 Αποτελέσματα Hazard Ratio από τρεις διαφορετικές πηγές για σύγκριση αποτελεσμάτων στην μελέτη IALT.**

Μελέτη: International Adjuvant Lung Cancer Trial (IALT)		
	Hazard Ratio	HR Διαστήματος εμπιστοσύνης 95%
Πηγή: Δημοσιευμένη Μελέτη IALT [29]	0,86	(0,76-0,98)
Πηγή: Δημοσιευμένη Μετα-ανάλυση [18]	0,91	(0,81-1,04)
Πηγή: Εφαρμογή	0,86	(0,73-0,98)

Η μελέτη ALP [30] που ξεκίνησε τον Ιανουάριο του 1994 διήρκησε 5.5 περίπου χρόνια. Ενώ αρχικά επιλέχθηκαν 1209 ασθενείς για να συμμετάσχουν ο τελικός αριθμός των συμμετεχόντων ήταν 1088 για λόγους διασφάλισης των αποτελεσμάτων εφόσον δεν πληρούσαν τα κριτήρια επιλογής. Από το σύνολο αυτών οι 540 ανήκαν στον πληθυσμό ελέγχου και οι 548 στον πληθυσμό υπό έρευνα με μέση παρακολούθηση και για τους δύο πληθυσμούς που έφτανε τους 64.5 μήνες. Το αποτέλεσμα της μελέτης είναι η καμπύλη επιβίωσης K-M της εικόνας 39 με ένα Hazard Ratio της τάξης του 0.96 (CI 95% 0,81-1,13). Με δεδομένα την εικόνα 39 τα χρονικά σημεία που ορίστηκαν από τον χρήστη, του πληθυσμού που συμμετείχαν στην μελέτη και τη μέγιστη και την ελάχιστη διάρκεια παρακολούθησης της μελέτης η εφαρμογή που έχει χρησιμοποιείται στην εργασία είχε σαν αποτέλεσμα τα στοιχεία του πίνακα 18 και ένα Hazard Ratio της τάξης του 0.97 που σημαίνει ότι δεν υπάρχει κάποια σημαντική διαφορά μεταξύ των 2 πληθυσμών (ελέγχου και υπό έρευνα) στη συνολική επιβίωση. Το Hazard Ratio που προέκυψε από την εφαρμογή συγκρίνεται με το αντίστοιχο από άλλες πηγές και συγκεκριμένα από τις δημοσιεύσεις της μελέτης ALPI στον πίνακα 19 .



Εικόνα 38 Καμπύλη KM από την τυχαιοποιημένη μελέτη Adjuvant Lung Project Italy (ALPI).



**Πίνακας 18 Αποτελέσματα εφαρμογής από την τυχαιοποιημένη μελέτη Adjuvant Lung Project Italy (ALPI).**

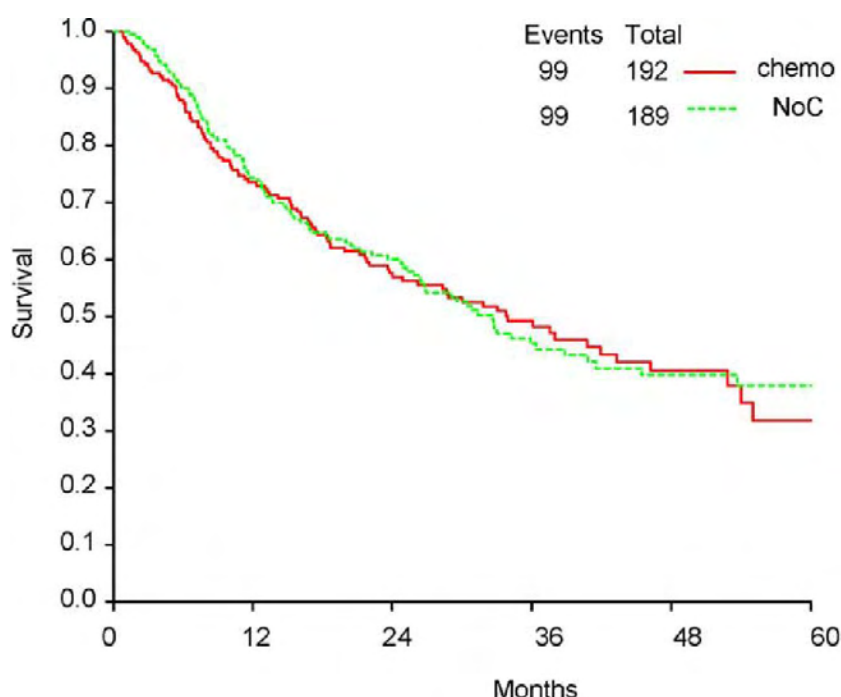
$t_i$	$S_{ci}(t_s)$	$S_{ri}(t_s)$	$R_{ci}(t_s)$	$R_{ri}(t_s)$	$C_{ci}(t)$	$C_{ri}(t)$	$D_{ci}(t)$	$D_{ri}(t)$	$\ln H_i(t)$	$Var_i$
0	100	100	540	548	0	0	18,42	25,15	0,30	0,09
5	96,59	95,41	521,58	522,85	0	0	30,58	39,52	0,25	0,05
10	90,93	88,20	491,00	483,33	0	0	42,58	46,71	0,11	0,04
15	83,04	79,67	426,00	414,79	22,42	21,83	34,41	23,70	-0,35	0,07
20	76,33	75,12	369,83	369,36	21,75	21,73	32,98	29,50	-0,11	0,06
25	69,53	69,12	315,80	318,61	21,05	21,24	26,49	18,51	-0,37	0,09
30	63,69	65,11	268,64	278,67	20,66	21,44	12,44	18,48	0,36	0,13
35	60,74	60,79	234,85	238,51	21,35	21,68	18,47	14,56	-0,25	0,11
40	55,97	57,08	194,74	201,55	21,64	22,39	10,81	11,92	0,06	0,17
45	52,86	53,70	160,94	165,93	22,99	23,70	10,07	3,25	-1,16	0,39
50	49,55	52,65	125,73	135,56	25,15	27,11	7,54	4,70	-0,55	0,33
55	46,58	50,82	94,55	104,68	23,64	26,17	2,63	2,80	-0,04	0,72
59	45,29	49,46	0	0	0	0	0	0	0,73	0,25

Τα αποτελέσματα του πίνακα 19 είναι πολύ κοντά μεταξύ τους είτε στο Hazard Ratio είτε στα διαστήματα εμπιστοσύνης. Το σημαντικό συμπέρασμα που προκύπτει είναι το γεγονός ότι δεν υπάρχει κάποιο όφελος στον πληθυσμό που χορηγήθηκαν φάρμακα (πληθυσμός υπό έρευνα) σε σχέση με τον πληθυσμό ελέγχου. Η εφαρμογή της εργασίας που λειτουργεί βασισμένη στις καμπύλες επιβίωσης K-M είναι πολύ κοντά στα πραγματικά αποτελέσματα από την δημοσίευση.

**Πίνακας 19 Αποτελέσματα Hazard Ratio από τρεις διαφορετικές πηγές για σύγκριση αποτελεσμάτων στην μελέτη ALPI.**

Μελέτη: International Adjuvant Lung Project Italy (ALPI)		
	Hazard Ratio	HR Διαστήματος εμπιστοσύνης 95%
Πηγή: Δημοσιευμένη Μελέτη ALPI [30]	0,96	(0,81-1,13)
Πηγή: Δημοσιευμένη Μετα-ανάλυση [18]	0,95	(0,81-1,12)
Πηγή: Εφαρμογή	0,97	(0,80-1,14)

Η μελέτη Big Lung Trial (BLT) [31] διήρκησε 6 χρόνια και συγκεκριμένα από το Νοέμβριο του 1995 μέχρι και τον Νοέμβριο του 2001 με σύνολο συμμετεχόντων 381 άτομα. Οι 192 από αυτούς ανήκαν στον πληθυσμό υπό έρευνα και οι υπόλοιποι 189 στον πληθυσμό ελέγχου με μέσο χρόνο παρακολούθησης μελέτης 34.6 μήνες. Τα αποτελέσματα από την μελέτη απεικονίζονται στην καμπύλη επιβίωσης K-M της εικόνας 40 που ουσιαστικά είναι η κατάσταση των ασθενών (ζωντανοί με το πέρασμα του χρόνου) που συμμετείχαν στην μελέτη. Το τελικό Hazard Ratio της συγκεκριμένης μελέτης είναι της τάξεως του 1.02, που πρακτικά σημαίνει δεν παρουσιάστηκε κάποια ουσιαστική διαφορά στην συνολική επιβίωση των δύο πληθυσμών που συμμετείχαν στην μελέτη. Με δεδομένα τα βασικά στοιχεία της μελέτης, την εικόνα 40 και τα χρονικά σημεία που ορίστηκαν από τον χρήστη η εφαρμογή της εργασίας είχε σαν αποτέλεσμα τα στοιχεία του πίνακα 20 και ένα Hazard Ratio της τάξης του 1.02 (CI 95 % 0,77-1,35). Η εφαρμογή συμφωνεί απόλυτα με το αποτέλεσμα από την δημοσίευση και αυτό φαίνεται καλύτερα και στον πίνακα 21 που συγκρίνονται τα αποτελέσματα από διαφορετικές πηγές.



Εικόνα 39 Καμπύλη KM από την τυχαιοποιημένη μελέτη Big Lung Trial (BLT).

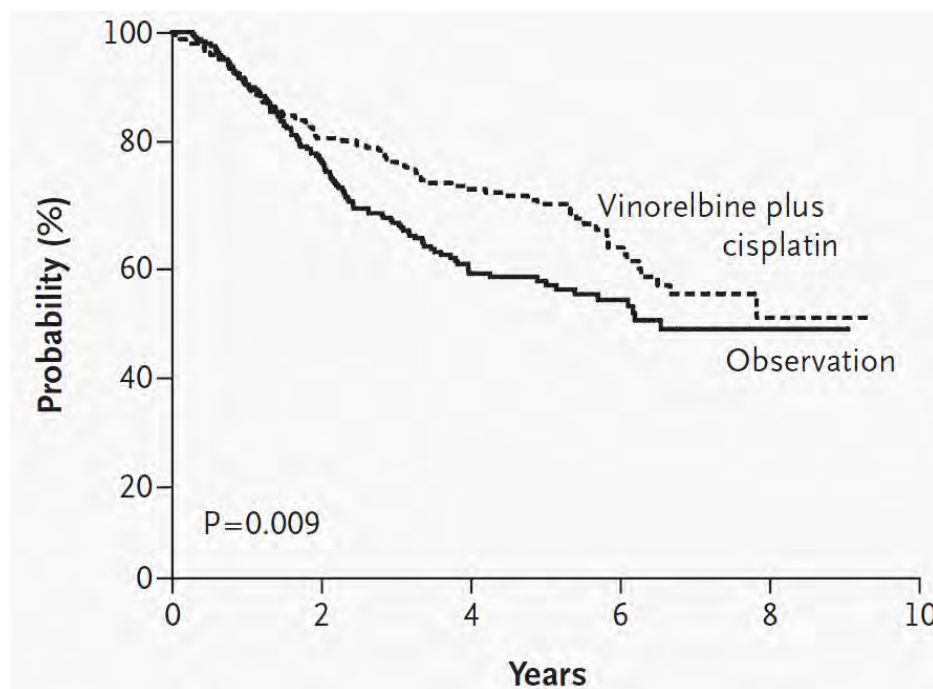
**Πίνακας 20 Αποτελέσματα εφαρμογής από την τυχαιοποιημένη μελέτη Big Lung Trial (BLT).**

$t_i$	$S_{ci}(t_s)$	$S_{ri}(t_s)$	$R_{ci}(t_s)$	$R_{ri}(t_s)$	$C_{ci}(t)$	$C_{ri}(t)$	$D_{ci}(t)$	$D_{ri}(t)$	$\ln H_i(t)$	$Var_i$
0	100,00	100,00	189,00	192,00	0	0	5,40	12,01	0,78	0,26
3	97,14	93,75	183,60	179,99	0	0	13,22	12,52	-0,03	0,14
6	90,15	87,23	170,39	167,47	0	0	30,98	26,13	-0,15	0,06
12	73,76	73,62	135,92	137,81	3,49	3,53	9,02	6,11	-0,40	0,26
15	68,87	70,35	121,34	125,92	5,57	5,78	9,96	15,40	0,40	0,15
20	63,21	61,75	106,02	105,21	5,35	5,31	6,83	8,79	0,26	0,24
25	59,14	56,59	93,92	91,30	5,28	5,13	9,91	6,19	-0,44	0,24
30	52,90	52,76	79,01	80,04	5,00	5,07	9,96	4,96	-0,71	0,28
35	46,23	49,49	59,72	64,94	9,33	10,15	6,80	9,69	0,27	0,22
45	40,97	42,10	44,10	46,04	8,82	9,21	3,22	6,45	0,65	0,42
54	37,97	36,20	35,20	34,09	5,68	5,50	0,14	4,08	3,39	7,20
59	37,82	31,87	0	0	0	0	0	0	-0,14	0,11
59	41,39	46,79	0	0	0	0	0	0	0,73	0,25

**21 Αποτελέσματα Hazard Ratio από τρεις διαφορετικές πηγές για σύγκριση αποτελεσμάτων στην μελέτη BLT.**

Μελέτη: Big Lung Trial (BLT)		
	Hazard Ratio	HR Διαστήματος εμπιστοσύνης 95%
Πηγή: Δημοσιευμένη Μελέτη BLT [31]	1,02	(0,77-1,35)
Πηγή: Δημοσιευμένη Μετα-ανάλυση [18]	0,95	(0,71-1,27)
Πηγή: Εφαρμογή	1,02	(0,79-1,26)

Η JBR.10 [32] είναι μια μελέτη που ξεκίνησε τον Ιούλιο του 1994 με τα τελευταία δεδομένα να καταγράφονται τον Απρίλη του 2004. Δηλαδή είναι μια μελέτη που έχει μια διάρκεια 10 σχεδόν έτη με 482 συνολικά ασθενείς να συμμετέχουν σε αυτή. Οι συμμετέχοντες χωρίστηκαν στον πληθυσμό ελέγχου και τον πληθυσμό υπό έρευνα σε 240 και 242 αντίστοιχα με μέση παρακολούθηση μελέτης 5.1 και 5.3 χρόνια αντίστοιχα. Η εικόνα 41 απεικονίζει 2 καμπύλες K-M που ουσιαστικά είναι η πορεία των ασθενών της μελέτης με βάση την θεραπεία που ακολουθούσαν. Το Hazard Ratio που προέκυψε από την συγκεκριμένη μελέτη είναι της τάξης του 0.69 που σημαίνει ότι υπάρχει μια διαφορά στην θεραπεία που χρησιμοποιήθηκε στον πληθυσμό υπό έρευνα έναντι της θεραπείας που χρησιμοποιήθηκε στον πληθυσμό ελέγχου. Δηλαδή είχαμε καλύτερα αποτελέσματα επιβίωσης σε όσους ασθενείς δοκίμασαν την νέα θεραπεία. Την εικόνα 41 και κάποια χρονικά σημεία που ορίστηκαν από τον χρήστη-όπως και στις προηγούμενες περιπτώσεις-χρησιμοποιήθηκαν σαν δεδομένα για να τεθεί σε λειτουργία η εφαρμογή της εργασίας που είχε σαν αποτέλεσμα ένα Hazard Ratio της τάξεως του 0.75 (CI 95% 0,52-0,91). Αυτός ο αριθμός προέκυψε από τα στοιχεία του πίνακα 22.



Εικόνα 40 Καμπύλη KM από την τυχαιοποιημένη μελέτη JBR.10.

**Πίνακας 22 Αποτελέσματα εφαρμογής από την τυχαιοποιημένη μελέτη JBR.10.**

$t_i$	$S_{ci}(t_s)$	$S_{ri}(t_s)$	$R_{ci}(t_s)$	$R_{ri}(t_s)$	$C_{ci}(t)$	$C_{ri}(t)$	$D_{ci}(t)$	$D_{ri}(t)$	$\ln H_i(t)$	$Var_i$
0	100	100	240	242	0	0	-0,45	4,20	2,23	-1,99
3	100,19	98,26	240,45	237,80	0	0	4,63	2,34	-0,67	0,64
6	98,26	97,30	235,82	235,46	0	0	9,01	6,29	-0,36	0,26
9	94,50	94,70	226,81	229,17	0	0	8,09	10,95	0,29	0,21
12	91,13	90,17	218,72	218,22	0	0	7,86	7,92	0,01	0,24
15	87,86	86,90	205,43	204,88	5,43	5,42	17,11	4,55	-1,32	0,27
20	80,54	84,97	185,25	197,06	3,07	3,27	8,42	8,04	-0,11	0,23
23	76,89	81,50	171,87	183,71	4,97	5,31	15,71	2,18	-2,04	0,51
28	69,86	80,53	151,52	176,12	4,65	5,40	6,07	3,80	-0,62	0,42
33	67,06	78,79	139,01	164,68	6,44	7,63	10,39	10,47	-0,16	0,18
40	62,05	73,78	119,69	143,50	8,93	10,71	11,53	4,15	-1,20	0,31
50	56,07	71,65	99,44	128,12	8,72	11,24	2,59	5,19	0,44	0,56
60	54,62	68,75	87,54	111,11	9,31	11,82	5,42	10,28	0,40	0,26
70	51,23	62,39	71,36	87,62	10,75	13,20	7,52	14,07	0,42	0,18
81	45,83	52,37	0	0	0	0	0	0		

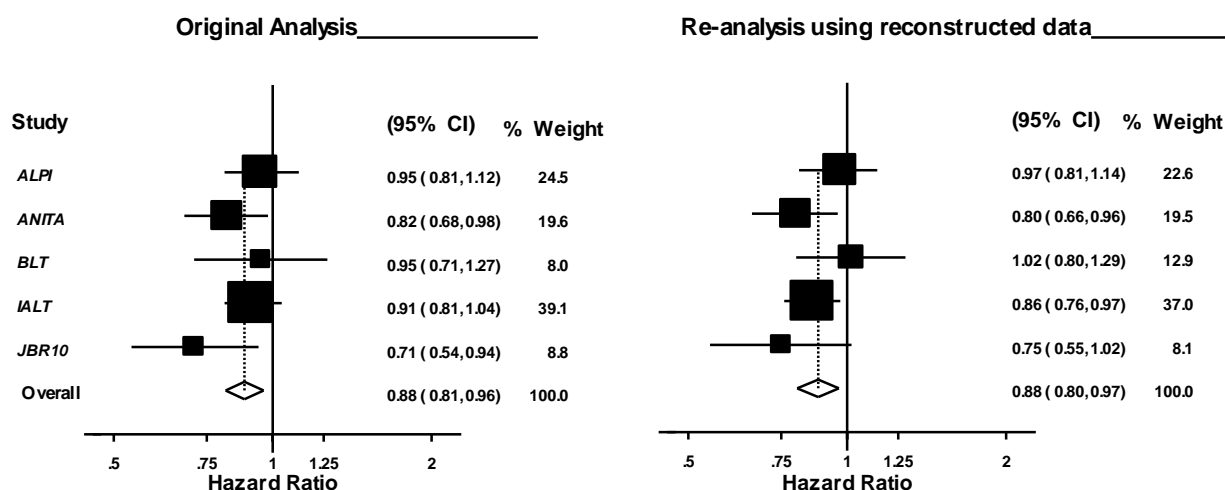
Η JBR.10 είναι η μοναδική μελέτη που παρατηρείται μια σημαντική διαφορά μεταξύ του Hazard Ratio της εφαρμογής που αναπτύχθηκε και των αποτελεσμάτων από την δημοσιευμένη μελέτη. Ο πίνακας 23 δείχνει τα αποτελέσματα του Hazard Ratio και τα διαστήματα εμπιστοσύνης που προκύπτουν από τρεις διαφορετικές πηγές για την μελέτη JBR.10.

**Πίνακας 23 Αποτελέσματα Hazard Ratio από τρεις διαφορετικές πηγές για σύγκριση αποτελεσμάτων στην μελέτη JBR.10**

Μελέτη: JBR.10		
	Hazard Ratio	HR Διαστήματος εμπιστοσύνης 95%
Πηγή: Δημοσιευμένη Μελέτη JBR.10 [32]	0,69	(0,52-0,91)
Πηγή: Δημοσιευμένη Μετα-ανάλυση [18]	0,71	(0,54-0,94)
Πηγή: Εφαρμογή	0,76	(0,51-1,01)

### 4.3.3 Μετα-ανάλυση

Οι συγκεκριμένες μελέτες συγκεντρώθηκαν και επεξεργάστηκαν έτσι ώστε να ελεγχθεί η αρχική υπόθεση «αν υπάρχει όφελος από την χρήση της σισπλατίνης κατά την διάρκεια της χημειοθεραπείας». Το αποτέλεσμα που προέκυψε από την σύνοψη των μελετών αυτών μέσω της μετα-ανάλυσης αντικατοπτρίζει την αποτελεσματικότητα της συγκεκριμένης θεραπείας. Οι παραπάνω μελέτες είχαν μέση παρακολούθηση μελέτης 5.2 έτη. Συνοψίζοντας αυτές της μελέτες μέσω της μετα-ανάλυσης προκύπτει ένα συνολικό Hazard Ratio της τάξης του 0.89 (CI 95% 0,82-0,96) [18]. Αυτό σημαίνει ότι υπάρχει ένα όφελος 11% για τους ασθενείς που ανήκαν στον πληθυσμό υπό παρακολούθηση δηλαδή τους ασθενείς που έκαναν χημειοθεραπεία που βασίζονταν στην σισπλατίνη. Δεν παρουσιάστηκε ετερογένεια στα αποτελέσματα των ασθενών που έκαναν χημειοθεραπεία στις παραπάνω μελέτες. Ο θετικός αντίκτυπος της χημειοθεραπείας ήταν υψηλότερο σε ασθενείς που είχαν γενικά καλύτερη κατάσταση και ανταπόκριση στην θεραπεία. Δεν υπήρχε καμία αλληλεπίδραση μεταξύ χημειοθεραπείας, φύλλου, ηλικίας, ιστολογίας, είδους εγχείρησης, προγραμματισμένης χημειοθεραπείας και συγκεκριμένης συνολικής δόσης της σισπλατίνης. Η εικόνα 42 είναι η σύγκριση των αποτελεσμάτων από τις δημοσιευμένες μελέτες με τα αποτελέσματα από την εφαρμογή όπως επίσης και το αποτέλεσμα της μετα-ανάλυσης που προέκυψε αυτές τις μελέτες σε σύγκριση με τα αποτελέσματα από την μετα-ανάλυση από την εφαρμογή.



Εικόνα 41 Σύγκριση μετα-ανάλυσης από δημοσιευμένο paper με τα αποτελέσματα της μετα-ανάλυσης από την εφαρμογή της παρούσας εργασίας.

**Πίνακας 24 Συγκριτικός πίνακας για τα αποτελέσματα της μετα-ανάλυσης όπως προέκυψαν από την εφαρμογή της παρούσας εργασίας σε σχέση με τα δημοσιευμένα αποτελέσματα.**

	Hazard ratio		95%Διάστημα εμπιστοσύνης	
	Δημοσιευμένη μετα-ανάλυση	Εφαρμογή	Δημοσιευμένη μετα-ανάλυση	Εφαρμογή
ALPI	0,95	0,97	(0,81-1,12)	(0,80-1,14)
ANITA	0,82	0,80	(0,68-0,98)	(0,62-0,99)
BLT	0,95	1,02	(0,71-1,27)	(0,79-1,26)
IALT	0,91	0,86	(0,81-1,04)	(0,73-0,98)
JBR.10	0,71	0,76	(0,54-0,94)	(0,51-1,01)
Μετα-ανάλυση	0,89	0,89	(0,82-0,96)	(0,80-0,98)

Τα αποτελέσματα της εφαρμογής που αναπτύχθηκε στην συγκεκριμένη εργασία έχουν κάποιες μικρές διαφορές με τα αυτά του Hazard Ratio από τις μεμονωμένες μελέτες. Η σύγκριση των αποτελεσμάτων παρουσιάζεται στον πίνακα 24 και στην εικόνα 42 που παρουσιάζει σε forest plot τα αποτελέσματα της δημοσιευμένης μετα-ανάλυσης και αυτά της εφαρμογής. Το γεγονός όμως ότι το τελικό αποτέλεσμα του Hazard Ratio που προκύπτει από την εφαρμογή ταυτίζεται απόλυτα με αυτό που παρουσιάζεται στην δημοσιευμένη μελέτη είναι πολύ θετικό για την απόδειξη της ορθής λειτουργίας της εφαρμογής. Είναι σημαντικό να αναφερθεί ότι το αποτέλεσμα προέρχεται από fixed effect model αλλά η εφαρμογή έχει την δυνατότητα να υπολογίζει και ένα αποτέλεσμα random effect μέσω μιας μεθόδου που έχει προταθεί από την Normand.

## 5 Συμπεράσματα

Η μέθοδος που ενσωματώθηκε στην εφαρμογή που κατασκευάστηκε στην παρούσα εργασία με την χρήση διαφόρων αλγορίθμων και συναρτήσεων είναι πολύ σημαντική. Υπάρχουν διάφορες πηγές που περιέχουν μελέτες οι οποίες έχουν ήδη πραγματοποιηθεί και τα στοιχεία αυτών δεν αρκούν για την χρησιμοποίησή τους σε μετα-αναλύσεις που θα μπορούσαν να οδηγήσουν σε ενδιαφέροντα συμπεράσματα. Με την συγκεκριμένη μέθοδο υπάρχει η δυνατότητα εκμετάλλευσης πολλών μελετών που δεν έχουν στοιχεία παραπάνω από μια καμπύλη KM, το πλήθος των συμμετεχόντων, το αντικείμενο μελέτης κ.α. Τα στοιχεία αυτά δεν μπορούν να χρησιμοποιηθούν για μετα-ανάλυση με την εφαρμογή κλασικών μεθόδων μετα-ανάλυσης. Για το λόγο αυτό αναπτύχθηκε η μέθοδος υπολογισμού του hazard ratio και της διακύμανσης και η χρήση αυτών για μετα-ανάλυση από καμπύλες KM όπως παρουσιάστηκε στο κεφάλαιο 2.2.

Η λειτουργία εντοπισμού της καμπύλης με την εφαρμογή της παρούσας εργασίας έχει σαν αποτέλεσμα τις μεταβλητές που χρειαζόμαστε για την πραγματοποίηση της μετα-ανάλυσης. Η σύγκριση με τις πραγματικές τιμές αυτών των μεταβλητών παρουσιάζεται στο κεφάλαιο 4.2 όπου περιέχει όλους τους ελέγχους και τις δοκιμές για την αποτελεσματικότητα της εφαρμογής. Οι διαφορές που προκύπτουν είναι μικρές και δεν επηρεάζουν το τελικό αποτέλεσμα που είναι ο δείκτης Hazard Ratio. Αυτό φαίνεται επίσης και στα αποτελέσματα της μετα-ανάλυσης με την συγκεκριμένη εφαρμογή σε πραγματικά δεδομένα στο κεφάλαιο 4.3.1-2.

Η εφαρμογή έχει το πλεονέκτημα ότι αποθηκεύει τα αποτελέσματα της εκάστοτε μελέτης σε αρχείο excel για περαιτέρω ανάλυση. Δηλαδή με τα δεδομένα εξόδου να είναι αποθηκευμένα, δίνεται στον χρήστη η δυνατότητα αξιοποίησης αυτών των Individual Patients Data για τη χρήση των κλασικών μεθόδων μετα-ανάλυσης όπως παρουσιάστηκαν στο κεφάλαιο 2. Για να ελέγξουμε την αποτελεσματικότητα της μεθόδου μπορούμε επίσης να τα χρησιμοποιήσουμε τα αποτελέσματα από την εφαρμογή σαν στοιχεία εισόδου σε

λογισμικό ανάλυσης δεδομένων και μετα-ανάλυσης όπως το STATA ή το ειδικό εργαλείο για μετα-αναλύσεις CMA.

Μετά την σύγκριση και με παρόμοιες εφαρμογές όπως το digitize it που κυκλοφορούν ήδη στο διαδίκτυο είναι προφανές ότι τα αποτελέσματα και η διαδικασία εύρεσης αυτών όπως περιεγράφηκε στο κεφάλαιο 4 μπορεί να γίνει με μεγαλύτερη ευκολία και με πολύ καλό αποτέλεσμα. Μπορεί να χρησιμοποιηθεί για όλους τους τύπους των καμπυλών K-M, αποθηκεύει τα δεδομένα σε αρχείο excel για περαιτέρω επεξεργασία και είναι αρκετά εύχρηστο εργαλείο. Έχοντας αυτήν σαν πρώτη έκδοση με παρατηρήσεις και με σχόλια από τους χρήστες θα μπορούσε να δημιουργηθεί ένα πολύ καλό εργαλείο που θα βοηθάει τέτοιου είδους μελέτες.



## 6 Βιβλιογραφία

- [1] S. Normand, «Meta-analysis: formulating, evaluating, combining, and reporting,» *Stat Med*, τόμ. 18, pp. 321-359, 1999.
- [2] M. Borenstein, L. V. Hedges, J. P. T. Higgins και Hannah R. Rothstein, *Introduction to Meta-Analysis*, United Kingdom: John Wiley & Sons, 2009.
- [3] T. Trikalinos, G. Salanti, E. Zintzaras και J. Ioannidis, «Meta-analysis methods,» *Adv Genet*, τόμ. 60, pp. 311-334, 2008.
- [4] T. Chalmers, J. Berrier, H. Sacks, H. Levin, D. Reitman και R. Nagalingam, «Meta-analysis of clinical trials as a scientific discipline. II: Replicate variability and comparison of studies that agree and disagree,» *Stat Med*, τόμ. 6, pp. 733-734, 1987.
- [5] S. D. F., J. Berlin, S. Morton, I. Olkin, G. Williamson, D. Rennie, D. Moher, B. Becker, T. Sipe και S. Thacker, «Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology,» (*MOOSE*) group. *Jama*, τόμ. 283, pp. 2008-2012, 2000.
- [6] Sharon-T. Normand, «Meat-analyis: Formulating, Evaluating, Combining, and Reporting,» *Statistics in Medicine*, τόμ. 18, pp. 321-359, 1999.
- [7] A. Cipriani και J. Geddes, «Comparison of systematic and narrative reviews: the example of the atypical antipsychotics,» *Epidemiologia e Psichiatria Sociale*, τόμ. 3, αρ. 12, pp. 146-453, 2013.
- [8] B. J. 3rd, «The promise and problems of meta-analysis,» *The New England Journal of Medicine*, τόμ. 8, αρ. 337, pp. 559-561, 1997.
- [9] M. Egger, G. D. Smith και A. N. Phillips, «Meta-analysis-Principles and Procedures,» *BMJ*, τόμ. 315, pp. 1533-1537, 1997.
- [10] S. Sauerland και C. Seiler, «Role of systematic reviews and meta-analysis in evidence-based medicine,» *World Journal of Surgery*, τόμ. 29, αρ. 5, pp. 582-587, 2005.
- [11] «The Cochrane Collaboration,» 1993 . [Ηλεκτρονικό]. Available: <http://www.cochrane.org/>.
- [12] «BMJ,» 1988. [Ηλεκτρονικό]. Available: <http://www.bmj.com/>.
- [13] J. Maksimović, «The Application of Meta-Analysis in educational Research,» *FACTA UNIVERSITATIS*, τόμ. 10, αρ. 1, pp. 44-55, 2011.
- [14] M. R. Barrick και M. K. Mount, «The Big Five Personality Dimensions and Job Performance: a Meta-Analysis,» *Personel Psychology*, τόμ. 44, αρ. 1, 1991.
- [15] A. Andrews, I. Zinger, R. D. Hoge, J. Bonta, P. Gendreau και F. T. Cullen, «Does Correctional Treatment Work? A Clinically Relevant and Psychologically Informed Meta-Analysis,» *Criminology*, τόμ. 28, αρ. 3, pp. 369-404, 1990.
- [16] M. K. Parmar, V. Torri και L. Stewart, «Extracting Summary Statistics to Perform Meta-Analysis Of the Published Literature for Survival Endpoints,» *Statistics in Medicine*, τόμ. 17, αρ. 24, pp. 2815-2834, 1998.
- [17] D. Altman και J. Bland, «Time to event (survival) data,» *BMJ*, τόμ. 317, pp. 468-469, 1998.
- [18] K. Gurusamy, «<http://www.ucl.ac.uk/>,» [Ηλεκτρονικό]. Available: [http://www.ucl.ac.uk/surgiscience/prospective\\_students/programmes/msc\\_evidence\\_healthcare/tabbbed\\_box\\_ebh/Sample\\_3.pdf](http://www.ucl.ac.uk/surgiscience/prospective_students/programmes/msc_evidence_healthcare/tabbbed_box_ebh/Sample_3.pdf).
- [19] J.-P. Pignon, H. Tribodet, G. V. Scagliotti, J.-Y. Douillard, F. A. Shepherd, R. J. Stephens, A. Dunant, V. Torri, R. Rosell, L. Seymour, S. G. Spiro, E. Rolland, R. Fossati και D. Aubert, «Lung Adjuvant Cisplatin Evaluation: A Pooled Analysis,» *Journal of Clinical Oncology*, τόμ. 26, αρ. 21, 2008.

- [20] L. R. Arends, M. G. M. Hunink και T. Stijnen, «Meta-analysis of summary survival curve data,» *Statistics in Medicine*, τόμ. 27, αρ. 22, p. 4381–4396, 2008.
- [21] M. Simmonds, J. Tierney, J. Bowden και J. Higgins, «Meta-analysis of time-to-event data: a comparison of two-stage methods,» *Research Synthesis Methods*, τόμ. 2, pp. 139-149, 2011.
- [22] F. Siannis, J. Barrett, V. Farewell και J. Tierney, «One-stage parametric meta-analysis of time-to-event outcomes,» *Stat Med*, 2010.
- [23] S. Michiels, P. Piedbois, S. S. N. Burdett, L. Stewart και J. Pignon, «Meta-analysis when only the median survival times are known: a comparison with individual patient data results,» *Int J Technol Assess Health Care*, τόμ. 21, pp. 119-125, 2005.
- [24] P. Moodie, N. Nelson και G. Koch, «A non-parametric procedure for evaluating treatment effect in the meta-analysis of survival data,» *Stat Med*, τόμ. 23, pp. 1075-1093, 2004.
- [25] T. Pernerger, « Estimating the relative hazard by the ratio of logarithms of event-free proportions,» *Contemp Clin Trials*, τόμ. 29, pp. 762-766, 2008.
- [26] J. T. Rich, J. G. Neely, R. C. Paniello, C. C. Voelker, DPhil, B. Nussenbaum και E. W. Wang, «A practical guide to understanding Kaplan-Meier curves,» *Otolaryngology-Head and Neck Surgery Journal*, τόμ. 143, αρ. 3, pp. 331-336, 2010.
- [27] J. Ventre, L. Fine, B. Bell και A. Arbor, «A Programmer's Introduction to Survival Analysis Using Kaplan Meier,» σε *PharmaSUG 2011*, Nashville Tennessee, 2011.
- [28] D. Collett, *Modeling Survival Data in Medical Research*, Second Edition, Chapman & Hall/CRC, 2003.
- [29] J. G. Ibrahim, «Northeastern Illinois Chapter of the ASA - Summer 2005 Workshop,» 2005. [Ηλεκτρονικό]. Available: [http://www.amstat.org/chapters/northeasternillinois/pastevents/presentations/summer05\\_Ibrahim\\_J.pdf](http://www.amstat.org/chapters/northeasternillinois/pastevents/presentations/summer05_Ibrahim_J.pdf).
- [30] M. Sydes και J. Tierney, «Excel meta-analysis tool,» MRC Clinical Trial Unit, London, 2004.
- [31] T. H. Cormen, C. E. Leiserson, R. L. Rivest και C. Stein, *Introduction to Algorithms*, τόμ. 24, MIT Press, 2000.
- [32] «Matlab The language of Technical Computing,» [Ηλεκτρονικό]. Available: <http://www.mathworks.com/products/matlab/>.
- [33] «Graph Digitizer Software,» [Ηλεκτρονικό]. Available: <http://digitizeit.de/>.
- [34] J.-Y. Douillard, R. Rosell, M. D. Lena, F. Carpagnano, R. Ramlau, J. L. Gonzáles-Larriba, T. Grodzki, J. R. Pereira, A. L. Groumellec, V. Lorusso, C. Clary, A. J. Torres, J. Dahabreh και P.-J. So, «Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB–IIIA non-small cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial,» *The Lancet Oncology*, τόμ. 7, αρ. 9, p. 719–727, 2006.
- [35] R. Arriagada, B. Bergman, A. Dunant, J.-P. Pignon και J. Vansteenkiste, «Cisplatin-Based Adjuvant Chemotherapy,» *The New England Journal of Medicine*, τόμ. 350, 2004.
- [36] G. V. Scagliotti, R. Fossati, V. Torri, L. Crino, G. Giaccone, G. Silvano, M. Martelli, M. Clerici, F. Cignetti και M. Tonato, «Randomized Study of Adjuvant Chemotherapy for Completely Resected Stage I, II, or IIIA Non–Small-Cell Lung Cancer,» *Journal of the National Cancer Institute*, τόμ. 95, αρ. 19, 2003.
- [37] D. Waller, M. Peake, R. Stephens, N. Gower, R. Milroy, M. Parmar, R. Rudd και S. Spiro, «Chemotherapy for patients with non-small cell lung cancer: the surgical setting of the Big Lung Trial,» *European Journal of Cardio-thoracic Surgery*, τόμ. 26, αρ. 1, 2004.
- [38] T. Winton, R. Livingston, D. Johnson, J. Rigas, M. Johnston, C. Butts, Y. Cormier, G. Goss, R. Incullet, E. Vallieres, W. Fry, D. Bethune, J. Ayoub, K. Ding, L. Seymour, B. Graham και M.-S. Tsao, «Vinorelbine plus Cisplatin vs. Observation,» *The New England Journal of Medicine*, τόμ. 352, 2005.
- [39] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, 2011, p. 147.
- [40] T. M. Therneau και P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, Springer, 2000.

- [41] M. Shuttleworth, «Meta Analysis and Meta Studies,» 2009. [Ηλεκτρονικό]. Available: <http://explorable.com/meta-analysis>.
- [42] C. R. Canoa, F. A. Carrillatb και F. Jaramillo, «A meta-analysis of the Relationship Between Market Orientation and Business Performance: Evidence From Five Contrinents,» *International Journal of Research in Marketing*, τόμ. 21, αρ. 2, pp. 179-200, 2004.
- [43] M. PC, F. J, S. G, A. G και F.-Y. YT, «Treatment of ocular hypertension and open angle glaucoma: meta-analysis of randomised controlled trials,» *BMJ*, τόμ. 331, 2005.
- [44] M. Egger, G. D. Smith και A. N. Phillips, «Metaanalysis: Principles and procedures,» *BMJ*, τόμ. 315, 1997.
- [45] R. Kaushal και D. W. Bates, «The Clinical Pharmacist's Role in Preventing Adverse Drug Events,» Harvard Medical School, 2001.

